INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313-761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
English vowels: Their surface phonology and phonetic implementation in vernacular dialects

Veatch, Thomas Clark, Ph.D.
University of Pennsylvania, 1991
ENGLISH VOWELS:
THEIR SURFACE PHONOLOGY AND PHONETIC IMPLEMENTATION IN
VERNACULAR DIALECTS

THOMAS CLARK VEATCH

A DISSERTATION
in
LINGUISTICS

Presented to the Faculties of the University of Pennsylvania
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy.

1991

Mark Liberman
Supervisor of Dissertation

William Labov
Committee Member

Richard Janda
Committee Member

Anthony Kroch
Graduate Group Chairperson

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
© Copyright 1991
by
Thomas Clark Veatch
To my teachers.
Acknowledgements

Many thanks are due, to teachers, friends, loved ones.

My first and best teacher, my mother, taught me that I can do anything I want to do, and that the entire world belongs to me, and to her, and to you, too.

Mark Liberman's sincere curiosity and respect for others' ideas allowed my ideas and my interest in them to flourish and grow. When he asked me to teach a class on vowels, the lecture became the acoustics chapter. His encouragement led me into the formal phonological treatment of my subject matter. When he said, "What do you mean, Vowels?", my answer became the phonology chapter. His ability to see to the heart of things kept me on track. His tremendous sense of perspective made my problems feel manageable. With an amazing ability to find the right solution to every problem, he is much like Ganesh, the Lord of Obstacles. He made it all happen.

I thank Bill Labov, the most interesting person in the world. I owe my career to his influence. Most of the problems I find interesting are problems that he has introduced me to. In some way mysterious to me, when I worked for and with him, I learned to love my work. He also gives the most insightful, to-the-point, constructive, and pungent comments on written work I have ever encountered. I have always appreciated his respect for his students' intellectual independence and his often severe practice of respecting others by letting them find their way. A moment with Bill is an exciting and thought-provoking moment.

Rich Janda helped me translate my chapters from idiolectal Tom-ese into something that more phonologists may understand. His support, assistance, and friendship were invaluable.
Mitch Marcus, who with Aravid Joshi has been sugar daddy to the linguistics department, provided the working environment, access to computing facilities, and moral support necessary to learn the computational skills that made possible this dissertation, and indeed, this style of linguistic/phonetic research. He supported me at a time when it seemed that linguistics could not, because learning to use computers in a sophisticated way to do phonetics research seemed antithetical to doing linguistics. Without his support this dissertation would have been much different, and much worse.

Thanks also to the other professors in the Departments of Linguistics at the University of Pennsylvania and Stanford University, who contributed to my education as a linguist and a phonetician: Leigh Lisker, John Fought, Gillian Sankoff, Henry Hoenigswald, Don Ringe, Ellen Prince, Richard Larson, John Rickford, Ivan Sag, Tom Wasow, Bill Poser, Will Leben, Joan Bresnan, and Charles Ferguson.

Thanks to my excellent Hindi teachers, Surendra Gambhir and Santwana Nigam, whose careful tutelage was indeed useful to me (page 11).

Thanks to all my teachers, especially my high-school math teachers, who patiently taught me the two things I believe I truly understand: high school algebra and trigonometry.

I would like to thank my colleagues who contributed their data and insights to this work. Sharon Ash and William Labov are studying the perceptual consequences of sound change in Chicago and other American dialects in the project on Cross-Dialectal Comprehension. Peter Patrick is conducting a sociolinguistic study of a neighborhood in Kingston, Jamaica. Otto Santa Ana has completed a sociolinguistic survey and sociophonetic analysis of the Chicano community of Los Angeles, California. Crawford Feagin has studied Anniston, Alabama speech for many years. I have benefited over and over from their knowledge, their help, their encouragement, and their excellent fieldwork.

Thanks to Pamela Saunders, Irene Vogel, Akin Labi, Brian McHugh, Kenjiro Matsuda, Mary Jack, Brian Teaman, Charles-James N. Bailey, Crawford Feagin, Beth Ann Hockey, Bill Reynolds, Sherry Ash, and Laura Koenig, for their help with comments on various chapters.

My fellow students are among my best teachers. David Graff gave volumes of useful comments on each of my chapters. During my apprenticeship in Labov’s laboratory, we
would often face frustrating, intricate and apparently unsurmountable technical problems. When in the face of yet another I wished I could curl up and die, his eyes would twinkle and he would walk over to pull out a manual, muttering, “How interesting,” under his breath, and learn how to solve it. I hope that I have acquired even a fraction of his unshakeable curiosity. Sherry Ash kept the Linguistics Lab together and aided in numerous ways. Christine Zeller helped me in thinking about the basic issues in the phonology chapter. The students in the 1991 Introduction to Phonetics class taught me about teaching, gave me their work (page 13), and stimulated me to think about fundamental issues. Nicola Bessell was a fountain both of wisdom and of references to the phonological and phonetic literature.


Pamela Saunders, my favorite, has taught me many, many things. Without her, I would remain a child. Thanks, Pam.

Thanks to a little old Indian man, for teaching me how to string green beans: clean, fast, and with a good attitude. Thanks to a young Indian woman for telling me (not in so many words) to go back to graduate school.

Thanks are due for financial support to the National Science Foundation and the taxpayers of the United States for an NSF Graduate Fellowship, to the American Institute of Indian Studies and the taxpayers of India for an AIIS Hindi language training fellowship, to the Department of Linguistics at the University of Pennsylvania for a semester’s financial support, and to the Linguistic Society of America and the generosity of Charles Ferguson and Shirley Heath for the Language in the U.S.A tuition fellowship to the Linguistics Institute.
Abstract
Abstract

English Vowels: Their Surface Phonological Structure and Phonetic Implementation in Vernacular Dialects.

Author: Thomas Clark Veatch
Supervisor: Mark Yoffe Liberman

Does phonetic grammar exist? The system of phonetic implementation relates surface phonological structures to measurable phonetic forms. These aspects of the linguistic system is studied theoretically and in the vernacular speech of four related dialects. It is seen whether this phonetic system may have linguistic (i.e., language-particular) aspects, by investigating dialects which are quite different in phonetic details. A chapter explains the mapping from articulatory configurations to formant structure, and derives the basic facts of acoustic phonetics. Then the surface phonological structure of a useful fictional dialect, "Reference American," is explored, combining phonological underspecification, autosegmental theory, and privative feature theory in a formal, symmetrical, and simple representation of its surface vowel inventory. After further discussions of theoretical background and methods of measurement and phonological classification, four dialects are studied, including Jamaican Creole, Chicago White English, Alabama English, and Los Angeles Chicano English. These dialect studies investigate how vowels are structured, produced, reduced, and coarticulated in English dialects by describing surface regularities of phonetic performance. The inventory and surface phonological structure of the dialect is described. Large acoustical studies of vowel phonetics in unrestricted vernacular speech (taken from sociolinguistic interviews) are conducted (7 speakers, 16000+ measurements). I describe the average formant frequency patterns corresponding to each phonological category, in one case (Jamaican) by giving a formal system of rules for deriving the acoustical patterns from the phonological structures. General phonetic rules are applied in language-particular ways to generate the observed acoustical distributions. Vowel reduction as a function of phrasal stress is investigated, finding in most cases that these complexes of phonetic shifts are accurately described as shifts in the direction of a "reduction target". The reduction target appears to be different in different dialects. In the final chapter, the
effects of consonants on preceding vowel nuclei are shown to have quite different patterns in different dialects, which in some cases may be attributed to differences of phonological structure, but in others are to be understood as dialect-specific (therefore, linguistic) phonetic implementation patterns. Surface phonology and observable acoustic patterns are used as a window into the partly linguistic system of phonetic interpretation.
# Contents

Acknowledgements  iii  
Abstract  vi

1 Introduction  1

2 Acoustics. The Mapping from Articulation to Formant Structure  9
   2.1 Fundamental Facts About the Relation of Formants to Articulation  10
   2.2 Boundary Conditions for Standing Waves in the Vocal Tract  11
   2.3 $F_1:F_2:F_3=1:3:5$ in Uniform Tube & on Average in the Vocal Tract  12
      2.3.1 Rayleigh's Rule  14
      2.3.2 Explaining the Node-Antinode Rule  15
   2.4 Rayleigh's Rule Applied to the Mapping from Articulation to Acoustics  18
   2.5 Acoustic Rhoticity  23
   2.6 Transitions to Consonants  24
   2.7 Coronal Loci  26
   2.8 Implications for Phonetic and Phonological Theory  31

3 Phonological Preliminaries: The Vowel Structure of Reference American  36
   3.1 "General American English"  37
   3.2 English Lexical Sets  42
   3.3 Temporal Vowel Structure in English  47
      3.3.1 Degrees of Phonological Vowel Length  48
      3.3.2 Length and Complexity in Structuralist Treatments of English Vowels  53

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.3</td>
<td>Is /r/ a Glide?</td>
<td>55</td>
</tr>
<tr>
<td>3.3.4</td>
<td>The Vowel Within Syllable Structure</td>
<td>61</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Syllable Weight</td>
<td>64</td>
</tr>
<tr>
<td>3.3.6</td>
<td>Implications of the Glide Slot for Dialectal Variation and Historical Change</td>
<td>65</td>
</tr>
<tr>
<td>3.4</td>
<td>Static Vowel Structure</td>
<td>70</td>
</tr>
<tr>
<td>3.4.1</td>
<td>The Neutral Position</td>
<td>70</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Vowel Height as Multi-Dimensional</td>
<td>71</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Vowel Height as a Single, Numbered Dimension</td>
<td>74</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Privative Height Features</td>
<td>76</td>
</tr>
<tr>
<td>3.5</td>
<td>Reference American Vowel Structure</td>
<td>78</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Apparent Problems in the Base-6 System</td>
<td>81</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Features in Reference American</td>
<td>87</td>
</tr>
<tr>
<td>3.6</td>
<td>Phonetic Implementation of Reference American</td>
<td>96</td>
</tr>
<tr>
<td>3.7</td>
<td>Summary</td>
<td>102</td>
</tr>
<tr>
<td>4</td>
<td>Theoretical Background Issues</td>
<td>104</td>
</tr>
<tr>
<td>4.1</td>
<td>Linguistic Versus General Phonetics</td>
<td>104</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Types of Phonetic Facts</td>
<td>104</td>
</tr>
<tr>
<td>4.1.2</td>
<td>The Overlap of Phonetics and Linguistics</td>
<td>105</td>
</tr>
<tr>
<td>4.2</td>
<td>The Vernacular</td>
<td>106</td>
</tr>
<tr>
<td>4.3</td>
<td>Issues in Acoustic Phonetic Research</td>
<td>108</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Definitions</td>
<td>108</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Acoustic Phonetic Features</td>
<td>112</td>
</tr>
<tr>
<td>4.4</td>
<td>Phonetic Inference from Differences in Formant Frequencies Among Classes of Vowels</td>
<td>114</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Mismatch between Audible Differences and Differences in F1-F2</td>
<td>115</td>
</tr>
<tr>
<td>4.5</td>
<td>Lisker (1949) Replicated</td>
<td>118</td>
</tr>
<tr>
<td>4.6</td>
<td>Vowel Reduction</td>
<td>124</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Types of Vowel Reduction</td>
<td>125</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Phonetic Vowel Reduction</td>
<td>126</td>
</tr>
</tbody>
</table>
## 5 Methods

5.1 Phonological Methods .......................................................... 130
5.2 Data ..................................................................................... 131
5.3 Analysis Procedures .............................................................. 133
  5.3.1 Background ................................................................. 133
  5.3.2 A Twelve-Step Program for Acoustic Analysis ................. 135
5.4 Impressionistic Coding for Phrasal Stress ............................. 139
5.5 Clitics Versus Other Words ............................................... 143
5.6 Locating Acoustic Nuclei .................................................... 144
5.7 Exclusions, Deletions, Total Measurements. ...................... 145
5.8 Outlier Analysis ................................................................. 146
5.9 Statistical Analysis ............................................................. 150
  5.9.1 The Bootstrap .............................................................. 150
  5.9.2 Are the Means of Two Samples in 2-Space Significantly Different? 153

## 6 Jamaican Creole

6.1 Speaker Information ............................................................ 156
6.2 Surface Phonological Structure ........................................... 157
6.3 The Shape of Formant Space ............................................. 165
6.4 Impressions of Stressed Vowels ........................................ 168
6.5 Acoustic Correlates of Phonological Vowel Length .............. 168
  6.5.1 Duration Differences .................................................. 169
  6.5.2 Phonological Length and Vowel Quality ............................ 171
6.6 Mean Vowel Nuclei in JC .................................................... 173
  6.6.1 Peripherality .............................................................. 176
  6.6.2 An Acoustic-Phonetic Grammar for JC Nuclei ................ 177
6.7 Stress Effects .................................................................... 181
6.8 Summary ........................................................................... 187

## 7 Chicago White English

7.1 Characteristics of Speakers and Recordings ......................... 189
List of Tables

3.1 Wells’ Lexical Sets ................................................................. 45
3.2 Examples of some confusing lexical sets. .......................... 46
3.3 Contrasts between one versus two syllables. ....................... 50
3.4 Constraints on consonants after stressed vowels .............. 52
3.5 Trager & Bloch’s English Vowel Structure (1941) .......... 54
3.6 Numerical representation of height, #1. ............................. 75
3.7 Numerical representation of height, #2. ......................... 76
3.8 Preliminary Reference American vowel structure (base-5). 79
3.9 Low monophthongs in Reference American. .................. 79
3.10 Reference American vowel structure ................................. 81
3.11 Implicational constraints on the feature [round] ............. 89
3.12 Feature specifications for nuclei. ................................. 90
3.13 Feature structure of the Reference American vowel system 94
3.14 Impressionistic forms of vowels as derived from rules (1)-(7). 98
4.1 Same and Different in Perception and Acoustics. .............. 116
5.1 Confusion matrix of phrasal stress classifications by two coders. 142
5.2 Numbers of tokens, deletions, exclusions, measurements. .... 146
6.1 Jamaican Creole Vowel Structure (Base-6) .................... 158
6.2 Jamaican Creole Vowel Structure (Base-5) .................... 164
6.3 Jamaican mean vowel durations (ms) and short/long duration ratios (percent). 169
6.4 Universal Phonetic Implementation Rules ....................... 178
6.5 Linguistic Phonetic Implementation Rules for Jamaican Creole 179
6.6 Confusion matrix of phrasal stress classifications on two occasions. 182

xiv

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
List of Figures

2.1 Movement of air in a standing wave cyclically towards and away from around a velocity node ........................................................... 15
2.2 Nodes and anti-nodes of the uniform tube ........................................... 19
2.3 “Locus equations” for three places of articulation ............................... 29
3.1 Judy from Chicago: “...fires, 'n' wars...” ........................................... 52
3.2 The underlying representation of /or/. ................................................ 85
3.3 Possible glide structures ................................................................. 91
3.4 Phonological Rules and Phonetic Implementation Rules .................... 98
3.5 Locations of nuclei in the F1-F2 triangle generated by rules (1)-(6) .......... 99
4.1 Overlaid formant trajectories for pap and pep, separately..................... 121
4.2 Overlaid formant trajectories for pap and pep, together....................... 122
6.1 Lexical sets that are not distinguished in Jamaican Creole .................. 158
6.2 Derivation of ingliding for long mid vowels. ....................................... 160
6.3 2680 F1/F2 measurements ............................................................. 166
6.4 1673 F1/F2 measurements ............................................................. 167
6.5 Raw F1-F2 measurements of long and short vowel pairs ..................... 172
6.6 Juba: Bootstrapped mean distributions ............................................. 174
6.7 Roasta: Bootstrapped mean distributions .......................................... 175
6.8 Before and after rules 5-8 ............................................................... 179
6.9 JC Stress Effects (Juba): Stressed and unstressed vowels in non-clitic words only, and in both clitic and non-clitic words ............................... 184
6.10 JC Stress Effects (Roasta): Stressed and unstressed vowels in non-clitic words only, and in both clitic and non-clitic words ............................... 185
7.1 4470 F1/F2 measurements. Interviewed 1990. .............................................................. 198
7.2 1614 F1/F2 measurements. Interviewed 1988. .............................................................. 199
7.3 2328 F1/F2 measurements. Interviewed 1970. .............................................................. 200
7.4 Rita’s Vowels; Rita’s /E/ .....................................................................................................202
7.5 Bootstrapped mean distributions. .........................................................................................205
7.6 Bootstrapped mean distributions. ..........................................................................................206
7.7 Bootstrapped mean distributions. ..........................................................................................207
7.8 Arrows point from mean of stressed tokens to mean of unstressed tokens.
  ‘ means that only stressed tokens of that category occur ........................................... 213
7.9 Arrows point from mean of stressed tokens to mean of unstressed tokens.
  ‘ means that only stressed tokens of that category occur ........................................... 214
7.10 Arrows point from mean of stressed tokens to mean of unstressed tokens.
  ‘ means that only stressed tokens of that category occur ........................................... 215
8.1 1637 measurements of F1, F2 .................................................................................................233
8.2 Bootstrapped mean F1, F2 for each vowel class. .................................................................236
8.3 Effects of impressionistic phrasal stress on vowel nuclei in F1-F2 space. .........................241
9.1 1890 raw measurements of F1, F2. .......................................................................................254
9.2 Bootstrap estimates of mean F1, F2 for each vowel class ...................................................256
9.3 Phonetid Implementation Rules for Los Angeles Chicano English ..............................257
9.4 Stress effects in raw F1-F2 data, #1 ..................................................................................263
9.5 Stress effects in raw F1-F2 data, #2 ..................................................................................264
9.6 Effects of impressionistic stress in Los Angeles Chicano English .................................266
9.7 LOT vs. THOUGHT, with following consonants. ..............................................................271
10.1 Effects of following /l/ on vowel nuclei in four dialects: JC (Jamaican Creole), AE (Alabama English), LACE (Los Angeles Chicano English), and CWE (Chicago White English). ......................................................... 277
10.2 Effects of following /ng/ on vowel nuclei in four dialects: JC (Jamaican Creole), AE (Alabama English), LACE (Los Angeles Chicano English), and CWE (Chicago White English). ......................................................... 283
Chapter 1

Introduction

The system of phonetic implementation relates surface phonological structures to measurable phonetic forms. This aspect of the system of linguistic performance is studied in the vernacular speech of four related English dialects. This thesis considers whether this phonetic system may have some linguistic (that is, language-particular) aspects, by investigating dialects which are quite different in phonetic details. After describing the surface phonological structures which are the inputs to the phonetic implementation system, a number of important aspects of the phonetic structure (in terms of measurements of first and second formant frequencies) of the different dialects are characterized, including the average phonetic qualities of vowel nuclei, the effects of phrasal stress on vowels, and to a much lesser extent, the effects of following consonant environments. A number of surface regularities in these systems are described, and related to one another, in an investigation of the ways in which vowels are variably produced, reduced, and coarticulated in English dialects.

The first half of the dissertation is about general theoretical issues in phonetics and phonology. An understanding of the phonetics of English vowels must begin with principles of general phonetics. Without a fundamental understanding of what vowel sounds are as sounds and of how they are produced, a phonetic study of vowel production is uninterpretable. Chapter 2, The Mapping from Articulation to Formant Structure, provides

---

General phonetics is distinguished from linguistic phonetics below. In short, general phonetics is the study of the physical universals of speech production and perception; linguistic phonetics is the study of language-particular aspects of speech production and perception.
a fundamental understanding of these central issues in phonetics. As such it is a contribution to general phonetics, and may be read independently of the rest of the thesis, but to properly interpret the acoustic-phonetic studies in the body of the thesis it is necessary to have an understanding of the content of this chapter. Even those who understand a great deal about acoustic phonetics may be more convinced that the methods and results of the remainder of the thesis are valid and meaningful after considering the arguments in this chapter.

The phonetic substance of vocalic sounds, not just in English, but in all languages (and even vocalic sounds of produced by gorillas and baboons) must be understood from both articulatory and acoustic perspectives. Often one perspective or the other, but not both, is taken as fundamental, as exemplified in the shift from acoustics-based features for Jakobson, Fant, and Halle (1952) to articulation-based features for Chomsky and Halle (1968). Studies of formant frequencies undertaken by those who lean toward acoustics may be criticized or insufficiently understood by those who lean the other way; similarly, EMG studies of the activities of the articulators may be poorly understood by those who lean towards acoustics. This chapter explains the relationship between vocal-tract resonance frequencies and vocal-tract shapes, and thus attempts to bridge this gap. It shows how formant frequencies are derivable from configurations of the vocal-tract and how F1 and F2 are precise measures of mouth-opening, and of tongue-body frontness. It thereby helps to justify the importance of the studies of formant-frequency patterns which are taken up in the dialect chapters.

On the other hand, Chapter 2 may be taken as an independent contribution to a fundamental understanding of general phonetics. Though the ideas are quite old, their application to these issues is in many aspects new. It describes how the pattern of resonances of acoustic tubes like the vocal tract can be modified by changing the shape of the tube. The effects of constriction and widening on nodes and antinodes of standing waves in the tube are described. The theory is presented without use of mathematics (though the theory is mathematically expressible), and gives a method of reasoning simply about the effects of movements of the vocal tract on formant frequencies, and conversely, about how to determine what the vocal tract is doing, given a pattern of formant frequencies. The theory is qualitatively understandable, unlike quantitative theories in which equations
describing large numbers of concatenated tubes of varying cross-sectional area are solved in order to predict the resonant structure of the tube. The theory is shown to explain quite simply the fundamental empirical facts of the relationship between articulation and vowel acoustics, including some rather mysterious ones (e.g., the triply constricted articulation of [ʊ]). Also, an experiment is conducted and described which confirms one detailed prediction made about a previously unstudied (to my knowledge) acoustic pattern (the difference between retroflex and dental “locus” frequencies). Finally, it is pointed out that the acoustic dimensions that this theory gives prominence to correspond quite straightforwardly to the traditional phonological dimensions of height and frontness: formant frequencies F1 and F2 are shown to reflect the degree of mouth-opening and of tongue-body frontness, and both acoustic and articulatory dimensions may simultaneously be understood as the phonetic dimensions corresponding to phonological height and frontness.

The surface phonological structure of English is discussed in Chapter 3, Phonological Preliminaries. By surface phonology, I mean the structure of the sound system which is the output of the lexicon. The goal of this chapter is to characterize the phonological structures which are the input to the system of phonetic implementation that is the central focus of study here. For this reason, lexical- or morpho-phonology was not considered; the focus is on the true generalizations and symmetries that hold on the phonological surface. After preliminary discussions of the inventory of comparative sound-classes and of the incoherence of “General American English” as an object of phonological study, that chapter develops a phonological structure for a hypothetical dialect, “Reference American” (which conceptually replaces General American). The structure proposed there has two aspects: static and temporal. The static phonological structure is a 3 X 2 (“base-6”) system of nuclei distinguishing three heights and two degrees of frontness. The temporal, or sequential, structure distinguishes a nucleus and a post-nuclear glide within the syllable, where the glide slot is used to represent vowel length as well as high-front, high-back, and rhotic glides, /y, w, r/. Central to the proposed analysis is the argument that postvocalic /r/ is a glide. Movements of post-vocalic consonants into the glide slot (/r/, centuries ago, and /l/, contemporarily in some dialects) thus has considerable consequences for the vowel system and explains a number of historical and ongoing vowel mergers in English dialects. The evidence used in this chapter is that of complementary distribution and
phonetic similarity, rather than morphophonemic alternations.

Chapter 4 discusses various theoretical background issues. It makes a central conceptual distinction between linguistic and general phonetics. It characterizes "vernacular speech" and argues that this is historically and socially the most important form of language; it also is the form which contains the greatest variety of phonetic variation, which is of central interest in this thesis. Next, definitions are presented for the analytical unit, "acoustic vowel", etc., and the relations of phonological vowels and consonants to acoustic vowels and consonants are clarified. Criteria for including measurable phonetic features in the catalog of real phonetic features are suggested. Finally, the relationship of F1, F2 measurements to this catalog of features is discussed, and an argument is presented that significantly different F1, F2 measurements, which are larger than the difference limen, must reflect real phonetic differences (that are audible, given training) which are to be accounted for at some level by a theory of phonetic performance. Lisker (1949) is replicated.

Chapter 5, Methods, describes the analytical techniques used in describing each dialect's vowel system and its phonetic implementation. Some of these methods are relatively new and unknown, such as the technique for estimating inherent precision in statistics, called the bootstrap, while others are traditional linguistic methods (e.g., complementary distribution, contrast, and phonetic similarity). Statistical regularities in phonetic form are also adduced as evidence for structural phonological questions. Methods of impressionistic coding of phrasal stress are described, along with reliability tests.

In this way, the first half of the thesis establishes a phonological framework for English vowels and a phonetic theory for understanding the articulatory significance of acoustic measurements. It also clarifies some aspects of the theoretical context in which these measurements are to be understood and provides a set of methods which are likely to lead to interesting phonetic and phonological generalizations in the study of spoken vernacular dialects. The descriptive body of the thesis follows these initial chapters.

The study of vowels in the English languages is approached from a number of directions in this thesis. "English" is not taken to be a single object. There are many different dialects of English, which are often extremely different from one another in surface-phonological inventory, in phonetic realizations of corresponding sound-classes, and in the phonetic processes that partly constitute the performance of phonological forms.
In order to investigate differences and similarities in these aspects of speech performance, several dialects are studied: Chicago White English (CWE), Los Angeles Chicano English (LACE), Anniston (Alabama) English (AE), and (Kingston) Jamaican (mesolectal) Creole (JC). The geographical and social boundaries of these dialects are not established here; what was important for this study was simply that they be as different from one another as possible.

A number of reasons determined the choice of these four dialects. A major goal was to test the assumption of the universality of phonetic implementation. To maximize the phonetic differences and minimize the phonological differences, different but related dialects of a single parent language were examined, since as different "accents", they may be expected to share much of the same phonological system, but at the same time to be phonetically quite different. In fact, Bailey (1985) titles an appendix, "The underlying vowels of most English lects"; if most English dialects have the same underlying vowel system, then it may be inferred that abstract phonological differences are minimized by choosing English dialects. On the other hand in order to maximize the differences in phonetic characteristics, dialects of English were chosen that are as different from one another as possible.

A cue was taken from Labov's (1991) paper on the "Three Dialects of English." There he identifies Northern, Southern, and Low-Back Merged as the three main (phonologically defined) divisions among English dialects. Thus one dialect from each of these areas was chosen (Chicago, rural Alabama, and Los Angeles, respectively). Further, in case these three dialects turn out to be too similar to one another, a fourth dialect was chosen which was so radically different that it might be considered a separate, though closely related, language.

Another important problem in enhancing diversity in this data was to avoid the effects of linguistic conservatism. The more conservative or standardized the speech, the more similar the dialects might be. For this reason upper-class, highly-educated speakers were avoided, in favor of analyzing the more vernacular, more historically advanced speech of working-class speakers.

One problem with studying coarticulatory patterns in the laboratory is that many natural phonetic phenomena of this kind often disappear under laboratory conditions. This is
the Observer's Paradox: if you look at something very closely, the act of observing changes the thing observed, so that one can't tell what it would have been like if it hadn't been invasively observed. Formal, self-monitored speech has much less coarticulation and reduction, than natural, normal, informal and un-self-conscious conversation (cf. Labov 1986, Keating and Huffman 1984). Because this study attempts to document significant cross-dialectal differences in phonetic processes of reduction and coarticulation, is important to maximize the effects of phonetic processes and to maximize the geographical differences, and thus to use as data the most unmonitored, vernacular style of speech. The vernacular is important for other reasons (discussed on page 106); it is the natural, undistorted form of the language used in day-to-day communication among native speakers of the dialect. An excellent way to get relatively unmonitored, continuous vernacular speech is by means of the tape-recorded sociolinguistic interview, so the data comes from previous and ongoing sociolinguistic studies of communities that fit the above constraints.

The first part of each dialect study characterizes the surface phonological structure of the vowel system of that dialect. Each dialect's surface phonological (not phonetic) inventory is explored and represented within the phonological theory presented in Phonological Preliminaries. Without getting the phonology right, there is little hope of making sense of the phonetics.

The core of each dialect chapter is a set of acoustic-phonetic studies of vowel quality as found in the natural conversational speech of one, two, or three working-class native speakers of the dialect. Vowels are characterized by taking measurements of the first two formant frequencies, F1 and F2, at a representative point within the syllable. The second chapter justified the articulatory importance of F1, F2 measurements. Formant trajectories throughout the acoustic vowel\(^2\) as well as acoustic-vowel duration, were also measured, but are not discussed here.

Thousands of vowels are measured for each speaker, in order to bring to statistical significance the effects on most vowel classes of stress and of many of the consonants that occur in their environment. The easiest way to collect enough data to bear on a large number of these effects is to simply measure all the vowels that occur in a fairly long stretch of continuous speech. Some phonological classes occur very frequently, while others

\(^2\)For a definition of acoustic vowel, see page 109.
(like /oy/) are underrepresented, but the overall result is that many, many effects have sufficient data to make them statistically significant.

The effects of phonological class and of phrasal stress are explored for each dialect. Thus the core of the thesis may be seen as contributions to our knowledge of sound-change, of vowel reduction, and of coarticulation. The knowledge about coarticulation and reduction gained over the past thirty years in the laboratory\(^3\) is thus extended to natural speech data in different dialects. The knowledge gained by sociolinguistic research on sound change over twenty years\(^4\) is extended by investigating interactions of dialect diversity with stress and consonant effects, factors which have important roles in sound change.

Finally, each of the dialect chapters may also be seen as a contribution to the literature on the respective dialects; thus, to creole studies (I believe this is the first acoustic study of creole phonetics), to the literature on Southern States English, to the study of Northern Cities dialects, and to the study of linguistic diversity in California.

The concluding \(10^{th}\) chapter, returns to theoretical questions about phonetic grammar by comparing the effects of two particular consonant environments on vowels in the various dialects studied. Different effects of following laterals are explored, which might or might not be attributed to (redundant) phonological differences. The striking lowering effects of following /\(\eta\)/ on vowels in the Alabama speaker, as opposed to the more coarticulatorily natural fronting and raising effects of the same consonant on /i/, for example, in other dialects, are not plausibly attributed to differences in the phonological form of /\(\eta\)/ in different dialects, or to the phonological content of the vowels that are affected. Thus there appears to be a dialect-specific coarticulation rule in the phonetic implementation system, which accounts for the lowering of vowels before /\(\eta\)/ in this Alabama speaker's natural speech. It is a fact that “physically motivated” processes are usually not physically necessary, and can vary stylistically, so that the most natural, physically easiest and most simplified phonetic forms are restricted to certain styles; similarly, there are differences in these details across dialects. Thus “natural” processes, while physically sensible, are not physically necessary. “Hard” coarticulation does exist, since the tongue cannot move infinitely fast. But there is a large realm of “soft” coarticulation, which is not due to

\(^3\)Lindblom's (1963a) thesis is a landmark in this field.

\(^4\)This subfield was inaugurated by Labov, Yeager and Steiner's (1972) publication of A Quantitative Study of Sound Change in Progress, referred to below as LYS.
absolute physical constraints. Dialects can differ in patterns of soft coarticulation that may be physically easy but are not physically necessary. Speakers must learn these patterns because it is part of what differentiates one dialect or language from another. These “details of implementation” are a part of the knowledge of native speakers of particular dialects, a part of what they know as speakers of that dialect and not some other, which may be different. It therefore appears that the control of these rather intricate details is a part of the human language faculty.

Linguistic-phonetic grammar includes the learned rules of soft coarticulation and of prosodically governed (stress-related) vowel reduction. Through comparisons of some of these details in the dialects studied, the final chapter confirms the existence of these linguistic phonetic effects, and opens the way to further research on this intricate and interesting aspect of the linguistic system.
Chapter 2

Acoustics. The Mapping from Articulation to Formant Structure

This chapter may be read independently of the other chapters as a contribution to general\(^1\) phonetics. It presents a theory of the mapping from articulation to formant\(^2\) structure, and applies that theory to explain the main observed facts about that relationship. Within the perspective of this thesis, the purpose of this chapter is to lay the physical groundwork for the phonetic studies of English vowels that form the body of the thesis. The need here is to justify the use of measurements of the first two formant frequencies (F1 and F2) as representations of phonetic vowel quality. These measurements are in many ways superior to (more consistent, precise, and objective than) impressionistic characterizations, but are sometimes considered phonetically uninterpretable. This chapter shows, within a theory of the mapping from articulation to formant structure, how F1 and F2 are direct measures of the fundamental articulatory (and auditory, and phonological) dimensions of vowel quality: height and backness/roundness.

First, I outline the main observed facts about formant structure under various articulatory configurations. Then in the remainder of the chapter I describe an idealized model of the resonating vocal tract, as a lossless, uniform acoustic tube containing standing waves; I present results consistent with the hypothesis that the average vowel (for the speakers

---

\(^1\)For the distinction between general and linguistic phonetics, see page 104.

\(^2\)A formant is a resonance frequency of the vocal tract.
studied) is fairly well modelled in this way. Next, I describe and explain a theory of the variation in the frequency of these standing waves, derived from Fant (1960, 1968:216-217), Chiba (1941: Chapter 11), and originally Rayleigh (1894). The physical explanation is qualitative and includes no mathematics; it should be understandable to anyone who wishes to work through it. Finally, I apply that theory to explain the effects on formant frequencies of mouth opening, tongue-body frontness, and constrictions at bilabial, dental, velar, and retroflex places of articulation. I also apply the theory to predict the results of an investigation of acoustic correlates of voiced stops at three coronal places of articulation. The theory makes its most unusual predictions for sounds with multiple constrictions; these are found to be true. Finally, some consequences of the model for phonetic theory and for theories of phonological structure are explored.

### 2.1 Fundamental Facts About the Relation of Formants to Articulation

The main facts about formant frequency variation can be thought of as having to do with vowels on the one hand, and consonants on the other. Vowel height is inversely related to F1 frequency. That is, F1 for high vowels is relatively low in frequency, while F1 for low vowels is relatively high in frequency. Vowel frontness is related to F2 frequency: front vowels have higher-frequency F2, back vowels have lower-frequency F2. Consonantal constriction and closure also has effects on the formant frequencies. Closure at the lips results in falling F1, F2, F3 frequencies; coronal closure shows a “locus” effect, where F2 rises into coronal closure if it starts low, and falls into coronal closure if it starts high; fronted allophones of velar consonants (henceforth “front velars”, likewise “back velars”) closure results in an F2-F3 “pinch”, where F3 falls and F2 rises, not so for back velars; rhotic /r/ results in an especially low F3.

These fundamental, empirical observations of the acoustics of speech call out for explanation: Why do these relationships hold between articulation and acoustic structure? One traditional answer has been to point to a transmission-line model: the vocal tract is

---

³Chiba & Kajiyama state that “It was our original goal to write this book without making use of mathematical equations” (1941:i), though they do not. For accessibility's sake, this goal remains a worthy one.
conceptually sliced into a fairly large number (on the order of 40) of short tubes of particular cross-sections and lengths. Equations for deriving the resonance frequencies from this detailed model of the shapes of the many short tubes are computed, and the resonance structure falls out rather magically. Such "explanations" are intellectually dissatisfying; it is impossible, certainly for me, to comprehend the relations of 3 or more frequencies relative to a model of 40 cross-sectional areas and 40 tube-lengths. A computer can certainly understand it — or calculate its transfer function — but for a human being, the model is a black box and the relations between articulatory configurations and acoustic structure are opaque. We are unable to reason simply about the most basic aspects of our field: the effects of vocal-tract shape on sound. We need a model for the relations between articulation and acoustics, which is both quantitatively accurate and qualitatively interpretable — i.e., where the mapping is intuitively predictable from one level to the other. No such model has been applied to explaining all these relations. We will define, justify, and apply such a model in this chapter, explaining the fundamental facts of acoustic phonetics in an understandable way.

2.2 Boundary Conditions for Standing Waves in the Vocal Tract

Let us review a standard idealized model of the vocal tract. In this model, the vocal tract is represented by a lossless, uniform, 1-dimensional, acoustic tube. That is, the walls absorb no energy, the tube has a constant cross-sectional area along its length, and the air within the tube vibrates only in the direction of the length of the tube. The resonances of such a tube are identified with the standing waves that meet the boundary conditions of the endpoints of the tube.

Given a tube which is closed at one end and open at the other, what are the boundary conditions imposed on standing waves within the tube? The boundary condition imposed by the closed end of the tube is that air particles\(^4\) at the closed end cannot move back and forth. Because they are adjacent to a wall — the closed end — they may be compressed

\(^4\)Air particles are to be thought of as infinitesimal volumes of air rather than as atoms per se. Air volumes have a pressure, but atoms do not.
against the end of the tube, but they are not free to move. Thus pressure can fluctuate maximally at the closed end of an acoustic tube, but particle velocity is zero.

The boundary condition imposed at an open end of an acoustic tube is that air pressure must be equal to ambient air pressure. Air particles at the open end will respond to a pressure wave coming out of the tube not by compressing, since there is nothing to compress against; instead, they simply move back and forth. Thus at the open end of an acoustic tube, a standing wave has maximum fluctuation in volume velocity, but zero variation in air pressure.

In slightly different terminology, the velocity waveform\(^5\) has a node (a location of zero fluctuation) at the closed end, and an antinode (a location of maximum fluctuation) at the open end. A mnemonic for “node” is “no delta”, where delta is the mathematical symbol for “change”. In any period of a standing wave, there are two nodes and two antinodes alternating with each other.

The half-open (i.e., open at one end, closed at the other) tube, then, has a velocity node at the closed end, and a velocity antinode at the open end. The tube length represents, at the least, one quarter of the period of a standing wave; thus, the lowest resonance in a uniform acoustic tube is the quarter-wavelength standing wave. Other standing waves are also compatible with these boundary conditions. For example, the next higher-frequency standing wave has a node at the closed end, an antinode at the open end, and a node-antinode pair in between. Each successively higher-frequency standing wave has an additional node-antinode pair between the fixed endpoints. We will consider the first three of these, which in a uniform tube are resonances that may be called F1, F2, and F3.

2.3 \(F1:F2:F3=1:3:5\) in Uniform Tube & on Average in the Vocal Tract

The length of a uniform acoustic tube constitutes \(1/4\), \(3/4\), and \(5/4\) of the wavelength of the first three standing waves. If this length is \(L\), then the wavelengths of the three standing waves are \(L*4\), \(L*4/3\), and \(L*4/5\). Since frequency is inversely proportional to

\(^5\)By velocity waveform, I mean the waveform which describes the velocity of air along the length of the tube.
wavelength \((frequency = c/wavelength, \) where \(c\) is the speed of sound\), the three resonance frequencies are in the relation 1:3:5. It is for this reason that the neutral vowel, \([a]\), which has an articulatory configuration that approximates a uniform tube, has \(F_1, F_2, F_3\) formant frequencies of, for example, 500, 1500, 2500Hz.\(^6\)

If the formant structure of the uniform tube is also the average formant structure of the vocal tract, then we should find not just that \([a]\) has this pattern, but also that the means of \(F_1, F_2,\) and \(F_3\) should be in a 1:3:5 relation, and that the true mean of \(F_1\) can be estimated by dividing the mean of \(F_2\) by 3, or dividing the mean of \(F_3\) by 5.

A small experiment was carried out to test this hypothesis. The data used was a database constructed by an Introduction to Phonetics class. 11 students individually produced 15 tokens each of the 10 vowels \(i:/\) (as in beat), \(i/\) (as in bit), \(e:/\) (as in bait), \(e/\) (as in bet), \(æ/\) (as in bat), \(a/\) (as in pot), \(u/\) (as in bought), \(u:/\) (as in boot), \(ɔ/\) (as in but), \(ʊ/\) (as in put), \(ɔ:/\) (as in boat), in the environment: “Say \{b,d,g\}_t again” nearly balanced across high \((n=4)\), mid \((n=4)\), and low \((n=3)\), front \((n=6)\) and back \((n=5)\); they also measured the first three formant frequencies for each token. For each student, I estimated the “true” mean \(F_1\) in the three ways mentioned: sample-mean \(F_1\), (sample-mean \(F_2\))/3, and (sample-mean \(F_3\))/5. To see how close these estimates are, I calculated the standard deviation of the three estimates for each speaker. These are given in the following sorted list, rounded to the nearest Hertz:

\[4 \ 12 \ 13 \ 17 \ 28 \ 28 \ 29 \ 35 \ 36 \ 51 \ 52\]

The results have a median of 28 Hz, and a range of 4 to 52 Hz. Thus the three estimates of neutral \(F_1\) are generally within about 30 Hz of their mean; this is on the order of the hand measurement error for a single token (cf., Labov, Yaeger, and Steiner, 1972:29,32, and Vol. II, Figures, p. 9.) I conclude that the frequency relations of \(F_1, F_2,\) and \(F_3\) are fairly close to the 1:3:5 ratios that are predicted by a model in which the formant structure of the average vowel is identified with that of the uniform, lossless acoustic tube. We may reasonably suppose that this average is the norm, and that actual vowels may be thought of as deviating from it in various directions to various degrees.

\(^{6}\)This assumes a vocal tract 17.1 cm in length, and a speed of sound, \(c\), of 343 meters/second — that is, in a nearly average male vocal tract in under normal atmospheric conditions.
2.3.1 Rayleigh’s Rule

The acoustic structure of resonant sounds within the vocal tract can be derived from the model of the vocal tract as a uniform tube with F1, F2, F3 identified as standing waves in the tube, plus a “rule of thumb” which shows how formant frequencies may vary from this norm according to the location and degree of constrictions distributed along the tube. This rule predates modern phonetics; it is much older than the spectrograph machine. Fant (1960) found it in Chiba (1941), but it was originally stated by Lord Rayleigh, about a century ago. The rule predicts in a detailed way many of the basic results of acoustic phonetics which were discovered much later, results which even today seem qualitatively mysterious. This acoustic rule has rarely7 been used to explain basic facts in acoustic phonetics, which as we will see in this chapter, it can do quite accurately and successfully.

The rule expresses the effect on standing wave frequencies of constrictions in an acoustic tube, relative to the nodes and antinodes of the standing wave concerned. Given a lossless, uniform acoustic tube and a particular standing wave resonance in it:

At a velocity node: If the tube is constricted in the vicinity of a node of the standing wave, the frequency of the standing wave rises. Conversely, if the tube is widened in the vicinity of a node, its frequency falls.

At an antinode: If the tube is constricted in the vicinity of an antinode of the standing wave, the frequency falls. Conversely, if the tube is widened in the vicinity of an antinode, the frequency rises.

These principles are worth memorizing. The effects at nodes are opposite to the effects at antinodes, and the effects of constriction and widening are opposite to each other. Thus all that must be remembered is that antinode constriction lowers standing-wave frequency, and the other three combinations can be derived from this one.

Fant does not explain these principles, but states that they follow from elementary considerations of electrical circuit theory (chapter 1.4). The explanation appears to rely on a decrease or increase in the speed of sound due to a constriction at the location of an antinode or node, respectively.

7Though see Ohala (1985) for its use in explaining the feature “flat.”
2.3.2 Explaining the Node-Antinode Rule.

Consider a velocity node in a standing wave somewhere in the middle of a uniform tube. On either side of the node is a velocity antinode; you can think of two plugs of air around the velocity antinodes, moving in towards the node, and out again. As the plugs move closer together and farther apart, the air pressure between them rises and falls. In the middle, at the velocity node, the motion of the two plugs towards each other cancels out, and pressure varies rather than particle velocity. Consider two plugs of air (symbolized | |) moving in and out relative to a velocity node, labelled N, in Figure 2.1.

The pressure in the middle is inversely proportional to the space between the plugs. The closer they get to the middle, the more they squeeze the air between them, the less the space between them, and the higher the pressure at the node in between them. At a certain point dependent on the amplitude of the standing wave, the pressure between them gets so high that it slows down the movement of the plugs towards the middle, eventually stops them, and then pushes them out again.

This is how a standing wave works: the plugs move towards each other, building up pressure between them and reducing pressure outside of them; as pressure builds between them, the resistance to their inward movement increases, so that they slow down, and
eventually stop. At this point the pressure between them is raised and the pressure outside them is lowered, so that the plugs are pushed outwards towards the low-pressure region. As they travel outward, the pressure in the middle falls, and the pressure outside rises; when the pressure outside is greater than the pressure inside, it starts to push them back towards the middle again. It is this process of alternation of pressures as air plugs at the velocity antinodes move alternately towards each other and away, which constitutes the standing wave.

Now suppose that the tube containing the standing wave becomes constricted at the velocity node. This means that there is a smaller volume of space between the two velocity-antinode plugs. What happens now when the two plugs move towards each other? Because they are pushing into a smaller space, the pressure between them will rise more quickly, and in response to the heightened pressure between them, they will slow down and begin to move in the opposite direction sooner. Similarly, as the plugs move outward there is less of a space to draw on, so the pressure between them falls more quickly. Thus the standing wave will vibrate in and out more quickly when there is less volume in the tube at the velocity node; that is, when the tube is constricted there.

Suppose conversely that the tube is widened at the node. Then the plugs at the velocity antinodes will be able to move farther towards each other before raising the pressure between them a given amount. The larger volume between them allows them to push farther inward, and the pressure between them builds up more slowly, given a fixed initial inward velocity. Similarly as the plugs move outward, the greater volume between them means that the pressure will not fall as quickly for a given amount of outward movement. Thus the vibration of the plugs will occur more slowly, which means that the standing wave will have a lower frequency.

What about constriction at the velocity antinodes? This is an even simpler problem. The situation is just like letting pressurized air out of any cavity. Getting a flat tire is a good example: if the hole is large, the tire will go flat right away; if the hole is tiny, the pressure will leak out more slowly. The larger the escape route for the pressurized air, the more rapidly the pressure will fall. If the tube is constricted outside the velocity node, then high pressure will take longer to fall, because less air will move out past the constriction in a given unit of time; conversely, low pressure will take longer to rise, because air can't
move in as quickly to equalize the pressure when it has to move through a smaller opening as through a larger opening.

If the tube is widened at the velocity antinode, then high pressure can drop more quickly, and low pressure can be equalized more quickly too. Thus widening at the velocity antinode results in more rapid pressure fluctuation; i.e., the standing wave vibrates at a higher frequency.

We have thus explained Rayleigh's node-antinode principles, by which constriction or widening at a node or antinode of a standing wave will raise or lower, or lower or raise, respectively, the frequency of the standing wave.

It is important to note that this was called by Fant a "rule of thumb", which accurately describes the effects of relatively small changes in the shape of an acoustic tube. The rule is not as clearly applicable when the constrictions in the vocal tract are not symmetrical with respect to the nodes and antinodes. Consider a velocity node between two velocity antinodes, at the instant that pressure is maximum at the node, as in Figure 2.1, time $t_5$. If equal constrictions are made at the antinodes on both sides of the node, then the air that escapes from the cavity between the constrictions during the outward-moving phase of the standing wave will escape symmetrically out the sides. Half of the escaping air will leave on each side. Thus the symmetry of the constrictions around the node ensures that the air on both sides moves away from the node symmetrically. Similarly, if there is relatively low pressure at the location labelled N in Figure 2.1, as at time $t_1$, the incoming air will come in symmetrically if the constrictions are symmetric around the node.

Now consider what happens if a severe constriction occurs on the left of the location labelled N, but only a mild constriction occurs on the right: then, when the pressure is maximum, the air can more easily escape through the relatively unconstricted side. Because the escape routes for the release of the relatively high pressure in the cavity are not symmetrical, more than half of the air will move in the direction of the open side. Therefore, even the air at the location labelled N will move away from the greater constriction. Only the air closest to the severely constricted end moves toward that side. Thus the node, which was formerly at the location labelled N, is not there any more, because the air at that location moves out (and in) towards the more open side. The node itself has shifted towards the constriction.
In this way, when constrictions in the vocal tract are not symmetrical with respect to the nodes and antinodes, the nodes and antinodes move away from the positions that they occupy in the uniform tube. The important point to remember is that if the constrictions are symmetrical with respect to the nodes and antinodes, then they will remain in the same location as the tube shape changes.

The main reason, in fact, that the rule applies so well to large as well as small changes in the shape of the acoustic tube formed by the vocal tract is, as we will see below, the striking fact that constrictions are typically symmetric with respect to nodes and antinodes in the shaping of the vowel-like sounds of language. In particular, a number of vocalic sounds have been found to be distinguished by differences in the higher formants (particularly F2 and also F3). These are just standing waves with multiple nodes and antinodes, and these sounds generally do tend to have multiple constrictions.

2.4 Rayleigh's Rule Applied to the Mapping from Articulation to Acoustics

Next we will use these principles to determine the effect on F1, F2, F3, etc., of constrictions at varying points in the acoustic tube formed by the vocal tract. Remember the first three standing waves in the ideal half-open acoustic tube: these standing waves constitute F1, F2, and F3 in the model, and their frequencies are related as 1:3:5.

Consider the locations of the velocity nodes and antinodes of these three standing waves in a uniform acoustic tube (the nodes and antinodes are numbered uniquely for later convenience of reference):

The boundary conditions, of zero velocity fluctuation at the closed end, and of zero pressure fluctuation at the open end, are signified by locating a node and antinode at the respective ends of the tube. Each higher-frequency standing wave has one additional node-antinode pair equally spaced along the length of the tube.

I will repeatedly make the analogy from this acoustic tube to the vocal tract. If the vocal tract in the position of [a] is taken as a uniform acoustic tube, then the lips are at the open end, the glottis is at the closed end, the alveolar ridge is between A2 and N3, the palate or front velum is located around N3, and A3 is around the upper pharynx.
Figure 2.2: Nodes and anti-nodes of the uniform tube

F1: 1/4 wavelength standing wave.

_______________________________
| A1 | N1 |

F2: 3/4 wavelength standing wave.

_______________________________
| A2 | N3 | A3 | N2 |

F3: 5/4 wavelength standing wave.

_______________________________
| A4 | N5 | A6 | N6 | A5 | N4 |
Consider first a simple case: a constriction at the open end of the tube. The result of this constriction, which is at an antinode for each of the three resonances, is the lowering of all three frequencies. Lip-rounding, and the movement towards labial closure that is associated with transitions into labial consonants, creates a localized constriction at the open end of the vocal tract. This predicts the well-established acoustic effect: In transitions from vowels to labial consonants, all the formants typically fall in frequency.

What if the acoustic tube is constricted more globally, so that the degree of constriction is not local to just the opening (lips), for example, but gradually increases along the length of the tube? The effect on the higher wavelength standing waves (F2, F3, F4, ...) will have a relatively null, washed-out effect: the distributed, gradual constriction lowers the resonance frequencies by constricting at antinodes, and also raises them by constricting at nodes. These opposing effects largely cancel.

This fact is quite general: if the constriction is not local to one node or antinode, then it will wash out. One can consider the overall Rayleigh's-rule effect as the sum of constrictions:

**Relation 1:** The frequency of a standing wave is proportional to the sum of constrictions positively weighted as they are closer to a node, and negatively weighted as they are closer to an antinode.8

Non-local constrictions for a given standing wave will add to both the node and antinode sums, and thereby cancel. Thus the higher formants will not be significantly affected by constrictions that are relatively long (that is, distributed along the length of the vocal tract). For this reason, for example, F3 is less well correlated with vowel quality than F1 or F2: tongue-body constrictions that affect vowel quality are not particularly localized. Thus theories of the exact location and degree of constriction for vowels seem to be overspecified (cf., Woods' theory, discussed in the next chapter). A gradually increasing constriction

---

8This relation is expressed mathematically in terms of an area function of the vocal tract (which specifies how wide the tube is at each point relative to its average width) and a weighting function which weights constrictions positively at nodes and negatively at antinodes. The frequency change may be found by multiplying the values of the area function by the corresponding values of the weighting function at each point along the length of the tube, and adding (integrating) the results. The sum of positively weighted constrictions is balanced against the sum of negatively weighted constrictions, the larger one determines the direction of change. If the constrictions are positively weighted, and the widenings (negative constrictions) are negatively weighted, the effect is an even more positive one, and the frequency rises a great deal.
along the length of the tube will therefore primarily affect the 1/4-wavelength standing wave, namely, F1, because it has no node within the tube by which the effect would be washed out. Thus the prediction is that the frequency of the 1/4-wavelength resonance, namely F1, will fall due to a constriction at its antinode. The other resonances will be much less affected since the effects of simultaneous constriction at both nodes and antinodes which result from a non-localized constriction will cancel each other out.

Conversely, if there is a gradual opening of the vocal tract, which reaches maximum opening at the antinode at the open end, the frequency of the 1/4 wavelength wave will rise, and the other resonances will be relatively unaffected. The jaw, of course, has precisely this effect: as the jaw opens and closes, the acoustic tube formed by the vocal tract opens like a horn, and closes like a bottle. Note that jaw opening does not merely open the front of the mouth, it also moves the tongue body back into the pharynx, because it is hinged on a point above and behind the curved tube formed by the vocal tract. Thus the result of jaw opening is that the vocal tract opens at the open end and closes at the closed end, resulting in an extremely gradual, articulatorily distributed effect, even more than would result solely from front-of-mouth opening. This explains what may be the main fact of acoustic phonetics: as the mouth opens and closes, F1 rises and falls in frequency. F1 measures vocal-tract openness.

Let us now consider the effects of constrictions which are rather more localized along the length of the tube. If there is a constriction at the medial antinode of the second resonance, i.e., at 1/3 of the distance from the closed end to the open end, the second resonance frequency will fall. If the constriction occurs at the node of the second resonance, namely 2/3 of the distance from the closed end, the second formant frequency will rise. These predictions are born out by observing of the association of F2 frequency with tongue body frontness. The location of the tongue body is roughly between 1/3 and 2/3 of the way from glottis to lips.

If the tongue body is closer to A3 than N3, forming a greater constriction at the antinode location than at the node location, then F2 will be lower in frequency than the uniform-open-tube F2 frequency. In other words the tongue will be relatively back, making the vowel articulatorily back, and F2 will fall.
Conversely, consider how the constriction might be greater at the node than the antinode. The tongue body would be fronted towards the point 2/3 of the way from glottis to lips, making a relatively greater constriction at node than at antinode, thereby raising this resonance frequency, namely F2. Indeed, the 2/3 location is roughly at the front velum or palate, and front vowels as a result have a relatively high F2 frequency.

We have thus explained the second fundamental fact of acoustic phonetics: vowel frontness is precisely correlated with F2 frequency, where fronter vowels have higher F2 (closer to the node N3) and backer vowels, closer to antinode A3, have a lower F2. F2, to a first approximation, measures tongue-body position relative to A3 and N3, that is, within the middle third of the vocal tract.

We have here shown how to derive the frequencies of the first two formants directly from considerations of the shape of the acoustic tube. The main acoustic cues to vowel quality are direct correlates of the degree of mouth-opening and of the position of the tongue body relative to the node and antinode of the second resonance, that is, of the degree of tongue-body frontness.

These results show that F1 and F2 directly reflect articulatory configurations, and that the phonological and auditory dimensions of vowel space — height and frontness — are directly related to the two formant frequencies \textit{and} to the articulatory configurations that they reflect. The acoustic measurements of F1 and F2 that form the body of this thesis are to be understood in the light of these results. These measurements are not merely acoustic, or articulatorily and auditorily uninterpretable features derived from the signal. These resonance frequencies are direct reflections of articulatory mouth-opening and tongue-body frontness.

These rather global articulatory properties of the shape of the vocal tract are themselves the fundamental dimensions of variation of vowel quality. Thus auditory vowel height and frontness, F1 and F2, and articulatory degree of mouth-opening and of tongue-body frontness are three physically interconvertible representations of phonetic vowel quality. For this reason, F1-F2 space is itself an excellent representation of vowel quality (given a fixed vocal tract length).

This argument justifies the methods of this thesis, which extensively uses and interprets F1-F2 measurements in terms of height and frontness. Since such heavy reliance is put on
this argument, we will now consider how the theory it is based on performs with respect to the remaining main facts about the relationship between formants and articulation.

2.5 Acoustic Rhoticity

We will next consider F3, which is the 5/4-wavelength standing wave in the idealized model. F3 is known to be involved most saliently in the contrast between English /r/ and /l/ (in some dialects). Retroflex and lateral continuants may be distinguished on spectrograms\(^{9}\) in this way: [\text{\ae}] has a severely lowered F3, while [l] is often distinguished from it by a raised F3. While this is not true of all sounds labelled /r/ or /l/, this relationship is nonetheless a striking mystery of acoustic phonetics. How is F3 lowered so severely in rhotic sounds?

We may predict from the acoustic fact that there is a constriction at one or more antinodes of the 3rd resonance. The prediction is resoundingly true, as discussed in detail in Ohala (1985).\(^{10}\) It turns out that [\text{\ae}] has a constriction made either by the tip of the tongue curled far back, or by bunching and raising the tongue, so as apparently to approach the middle antinode, A6, of the 3rd resonance. Additionally, retroflex continuant /r/ in English occurs with a seemingly peculiar, localized constriction in the pharynx: “There is also a constriction in the pharynx below the epiglottis” (Lindau, 1975:27). If this occurs at the F3 antinode, A5, then the effect would be the observed one, a lowering of F3. Finally, it is well known that retroflex /r/ in English has lip-rounding. Thus not only is there a constriction at one antinode, there are constrictions at all three of the antinodes of the F3 standing wave. This is striking confirmation of the theory of the relations between acoustics and articulation.

Why should F3 be raised with a clear [l]? The theory predicts a constriction at nodes N5 and N6; contact of the tongue-tip behind the teeth during [l] would seem to support this prediction. Given the number of precise and accurate predictions made by Rayleigh’s node-antinode rule, we may have guessed that those [l]’s with raised F3 also have a localized dorsal constriction, at the location of node N6, as it appears they do.

The concomitant rounding of the lips which accompanies back vowels is quite similar

\(^{9}\)For example, V. Zue makes this point in spectrogram-reading instructions.

\(^{10}\)Thanks to Dave Graff for referring me to this paper.
to the simultaneous multiple articulations of [ɔ] and [i]. To lower F2, not only can there be a constriction at A3, there can also be a constriction at A2, namely at the lips. It is no coincidence therefore that back vowels, which have a lowered F2, should in most cases also have lip-rounding. Both articulations amount to antinode constrictions, and they have the same effect on F2. With both backing and rhoticity, completely unrelated articulations cooccur; the constrictions happen to be those which have the identical acoustic effect. This suggests that the acoustic effect is primary, and the articulation simply follows from the desired effect. A constriction at both antinodes results in a lower F2. The same holds for F3. Even when the constrictions are far apart, their effect is the same, and if the articulation simply follows from the desired acoustic effect, the separate constrictions that have no apparent articulatory connection make sense. Without the acoustics to make sense of the articulations, the cooccurrence of these independent constrictions is a mystery.

2.6 Transitions to Consonants

We may use the same principles to derive the effects on formant frequencies of closure associated with various consonantal places of articulation. Labial closure, as noted above, amounts to a local constriction at the open end of the acoustic tube. Since all the resonances have a node at this point, constriction at the lips lowers all the resonance frequencies.

"Velar" closure occurs over a range of places in the vocal tract, often classified into front and back velars. The front velars occur in my speech adjacent to front vowels; back ones next to back vowels. It is likely, though perhaps not proven, that this coarticulatory effect is a continuous one, not the result of a categorical division of velar consonants into front and back allophones. However, as a general rule, the front velars do have a particular acoustic structure which is shared by all of them, and not shared by back velars, namely the characteristic "pinch" of F2 and F3, where they move towards each other, finally almost merging just at the point of closure. Consider the effects of a constriction at or around node N3 and antinode A6, i.e., at the back of the palate or front of the velum. The locations of N3 and A6 are very close to one another, and a somewhat distributed closure such as may be made by the tongue dorsum will have a constricting effect on both. This would result in lowering the A6 resonance, namely F3, and raising the N3 resonance, F2. This is
in fact the commonly observed acoustic effect of front-velar closure.

Back velars have a constriction away from this location, in which an F3 antinode and an F2 node occur quite close to each other. As one might expect, they do not show this pattern of F2-F3 approximation.

Alveolar closure occurs between the F2 antinode at the lips and the F2 node at the palate. It thus contributes to both sums in Relation 1, above. The effect is therefore a balancing one. Apical closure is partly like node constriction, and partly like antinode constriction. If there is already node constriction (i.e., high F2), apical closure adds to both node constriction and antinode constriction; the result, on balance, is greater relative antinode constriction, and according to Rayleigh's rule, a lowering in F2. If there is already antinode constriction, i.e., lowered F2, then apical closure will have the opposite balancing effect, resulting in greater relative node constriction, and a rise in F2. Note that the balancing effect must be partial rather than complete; a low F2 will be raised, and a high F2 will be lowered, but the raising and lowering will not reach the midpoint.

This pattern is precisely the observed pattern of F2 changes when apical closure occurs; the "locus theory" for coronal consonants describes this effect. If a vowel with high F2 is adjacent to an apical consonant, then the transition into the consonant lowers F2; conversely a low vocalic F2 will rise, in the transition into an apical consonant. Even the detail that the "locus" is virtual rather than fully realized is explained: the relative imbalance between the node and antinode sums in Relation 1, above, is evened out by adding constants to both, but it is not completely restored. Therefore the transition to an apical consonant will move in the direction of the balance point, but will only move a fraction of the distance toward it, according to the size of the node and antinode sums in Relation 1, relative to the contribution of the apical constriction to each one.

The coronal "locus" has often been claimed, without apparent understanding, to reflect some deeper articulatory reality, so that there is some underlying resonance at the never-attained locus frequency that is characteristic of an actual articulatory state. But we see that this effect is entirely an artifact of the node-antinode balancing due to a constriction between a node and an antinode. No cavity has been isolated which resonates at this hidden frequency. Here we see that this deeper articulatory reality does not exist. There is no locus.
Rayleigh's rule can thus be applied to explain the finest details of formant structure, not just for vowels, but for transitions to consonants also. Its explanation of the locus theory must be considered an important step for the science of acoustic phonetics.

2.7 Coronal Loci

So far Rayleigh's rule has been applied to the explanation of established facts. Let us use it now to predict new facts, which to my knowledge have not previously been observed. Consider the effect of articulating an apical obstruct at various locations between the antinode at the lips and the F2 node around or behind the palate. By the theory, the farther back the constriction, the greater the effect of F2 node constriction relative to the effect of antinode constriction, and thus the higher the balance point, or locus frequency, of the second formant. Varying the constriction location in this way in a uniform tube corresponds to the articulatory differences between retroflex and alveolar (backer) vs. dental (fronter) obstructant locations.

An experiment was carried out to explore the different effects of various apical places of articulation on the formant transitions into the consonant. The measurements are similar to those of Sussman (1991).

Data and Method:

I produced voiced dental, alveolar, and retroflex stops in repetitive, equally stressed, CV syllables, 4 to an utterance, varying the vowel among /iy, ey, i, e, æ, o, ow, uw, u, a/. One utterance, for example, was of the form "dadadada." Two sets of four repetitions of each CV combination were spoken and digitized (16 bits, 10kHz), resulting in 2 sets * 4 repetitions * 3 places * 11 vowels = 264 tokens. Two signal-processing operations were performed: an RMS energy contour was calculated, to be used in locating the measurement

---

Dental and retroflex sounds in these contexts are not part of my native phonological inventory; I am a native Californian English speaker raised in Thailand (1-4), Okinawa (4-10), and California (10-23), and my parents are also raised on the West Coast. However, two years of intensive Hindi study in the U.S. and in New Delhi provided some training and experience in producing them for the purpose of communication.
points; and the signals were formant-tracked to measure formant frequencies.\textsuperscript{12} The onset and nucleus time-points were located by means of a nearly automatic procedure. The onset and offset of the acoustic vowel in each CV syllable was automatically located using an energy threshold: the onset of the vowel is located at the upward threshold-crossing in the amplitude contour, and the offset is located at the downward threshold-crossing. The crucial choice of an effective threshold is made by examining the values of the RMS energy contour at the onsets and offsets of a number of syllables, and picking an RMS value intermediate between the values for bursts and closures and the values for vocalic segments. The occasional jitter back and forth across the threshold level was eliminated by a "sloppy-crossing" threshold algorithm, where a "true" threshold crossing is taken to be one where after crossing the threshold the signal stays on the other side for some fixed time. The exact threshold crossing location is linearly interpolated between frames, so that if frame \( t \) is just under the threshold, and frame \( t+1 \) is far above the threshold, the crossing location is closer to the center time of frame \( t \) than to that of frame \( t+1 \). Using this method of automatic segmentation not only is a labor saving device, but also results in more consistent segmentation. The same criterion is used consistently throughout, where in hand segmentation the method may vary slightly from the beginning to the ending of a lengthy measurement session.

The threshold-based onset and offset locations are examined by eye in relation to the waveform. The threshold that was chosen resulted in only 6 out of 264 tokens with even slightly unreasonable segmentation points; these were modified by hand.

Formant frequencies for the frame whose center time was within the centisecond immediately following the onset location were extracted from the formant-track data. The nucleus formant frequencies were taken from the frame halfway between vowel onset and offset locations.

The formant frequencies of onset and nucleus for the three categories, dental, alveolar,
and retroflex, are plotted against each other in charts displaying F2 transitions. A line was fitted to each of the three classes of tokens. The results are in Figure 2.3.

These graphs display transitions between the onset and the nucleus. They are similar to those found in Sussman (1991), following Lindblom (1963a). Let us consider how they may be interpreted.

If the onset is higher than the nucleus for a given token, that token will occur above the Y=X line. If there is no transition, and the onset frequency is identical to the nucleus frequency, then that token will occur on the Y=X line, at its nucleus frequency. If the onset is lower than the nucleus, then the token will be plotted below the Y=X line. The magnitude of the transition equals the distance above or below the line. If in all transitions for a given consonant class the onset were lower than the nucleus, independent of nucleus F2 frequency, then the tokens would be distributed below the Y=X line. Bilabial consonants, with an onset F2 lower than the nucleus F2, should be arrayed in this way.

In these graphs, tokens with low-frequency F2 nuclei occur above the line (onset is higher than nucleus), while those with high-frequency F2 nuclei occur below the line (onset is lower than the nucleus). The point at which the regression line intersects the Y=X line, called the “Y=X-intercept” in the chart, is interpretable as the classic Haskins “locus” frequency: above that nucleus F2 frequency, onsets are lower than nuclei, while below that frequency, onsets are higher than nuclei, and just at the Y=X-intercept, there is no transition. The slope of the line\(^\text{13}\) shows the degree of applicability of the locus theory to the data: If the line is horizontal (slope=0), then no matter what the nucleus, the onset would start at the same frequency. This would occur if there were a fully realized locus, from which all transitions begin. If the line has slope=1 (that is, it is parallel to the Y=X line), then all the onsets are a fixed direction and distance from the nucleus. E.g., if F2 is 30 Hz lower at the release of a bilabial consonant than in the nucleus, independent of the nucleus F2 frequency, the slope of the fitted line in the transition graph would be 1, and its vertical offset would be 30 Hz below the Y=X line. Consonants with a locus equation of slope between 0 and 1 have a “virtual locus,” which is not attained in cases where the nucleus F2 is distant from the locus frequency. The closer the slope is to 0, the more

\(^{13}\)If it is indeed a straight line, then slope = \((\max(\text{onsetF2})-\min(\text{onsetF2}))/(\max(\text{nucleusF2})-\min(\text{nucleusF2}))\).
Figure 2.3: “Locus equations” for three places of articulation

All Coronal F2 transitions

Dental F2 transitions

Alveolar F2 transitions

Retroflex F2 transitions

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
actual, and less virtual, the locus is. The closer the slope is to 1, the more the transitions follow a constant pattern, going a relatively fixed distance in a fixed direction. For slopes greater than one, the range of onset frequencies should be greater than the range of nucleus frequencies, resulting in a set of transitions forming a reversed fan that spreads out, going backwards from the vowel nucleus into the preceding consonant. Consonants exhibiting such a pattern would actually have an anti-locus, a frequency from which formants move away, in the transition from a vowel to a consonant. Thus, for example, if the nucleus is a little distant in frequency from the anti-locus, the onset is even farther away from it. This pattern would be a surprising finding that is completely incompatible with a locus theory of consonant-vowel coarticulation.

The three consonant classes examined here fall into two groups: alveolar and retroflex on the one hand and dental on the other. The difference between the two groups is exactly as predicted by the node-antinode theory: consonants with a constriction closer to the lips and farther from the palatte contribute more to the sum of antinode constrictions than to that of node constrictions, and should have lower onset transitions in general than consonants with a constriction farther back in the mouth, closer to the F2 node. Thus the \( Y = X \) Intercept, i.e., the locus frequency, for the dentals is lower than for the alveolars and retroflexes, by about 160Hz.

Another difference between the two classes is that the locus equation’s slope is smaller, (the line is closer to horizontal) for the retroflex and alveolar classes than for the dentals. By the above interpretation, this means that the locus for the retroflex and alveolar consonants is closer to being fully realized than for the dental consonants. Conversely, the dentals have F2 transitions that are more nearly constant in pattern across different F2 values. In short, there is less of a locus effect with dentals than with alveolar or retroflex consonants. This is to be expected if alveolar/retroflexes are closer to the midpoint between node and antinode than dentals, since the balancing, “locus” effect is lessened, and the constriction becomes more purely one kind of effect or the other. At the midpoint between node and antinode, a constriction will make a maximally balanced contribution to both node constriction and antinode constriction sums, and the node-antinode balancing, which constitutes the locus effect, is maximal.

Having interpreted the difference between the dentals and the other coronals studied
here, the question remains, Why are retroflex and alveolar consonants so similar? The slopes of the fitted lines are within 2 percent, and the locus frequencies \((Y=X\text{-intercepts})\) are identical to 3 decimals. This near-perfect identity is quite puzzling, since the retroflex consonants were felt proprioceptively to be articulated farther back in the mouth than the alveolars, and also they sound more retroflex. There may be other acoustic cues to distinguish them, such as \(F3\) differences.

This patterning does correlate interestingly with the fact that speakers of Hindi, which has both dental and retroflex stops, classify English alveolars with their retroflex stops rather than with their dentals. The retroflex production of English coronal stops is a salient part of an Indian English accent. Perhaps native Hindi speakers learning English classify alveolar stops as retroflex stops on the basis of this acoustic similarity.

However, the identity of alveolar and retroflex stops is not predicted by the node-antinode theory, unless the node is located between the alveolar and retroflex places of articulation. Since the node is \(1/3\) of the distance from lips to glottis, this seems unlikely. Perhaps constrictions elsewhere in the vocal tract have counterbalanced the \(F2\)-raising effect of moving the constriction back, towards the node.

### 2.8 Implications for Phonetic and Phonological Theory

In this chapter, we have explained Rayleigh’s rule, that constriction at a standing-wave antinode lowers its frequency, vice versa at nodes rather than antinodes, and vice versa for widening rather than constriction. Viewing the resonating vocal tract as a half-open, ideal acoustic tube which normally (in the sense of “on average”) has a uniform cross-section, we applied this rule to relate many of the central observed facts of acoustic phonetics to particular (and correct) tube configurations, and to map them directly to articulatory configurations. Through this model we related the acoustic and articulatory effects of vowel height and frontness, rhoticity and laterality, and several effects of increasing constriction in the transitions from vowels to various consonants. We have shown that \(F1\) measures mouth opening, \(F2\) measures relative constriction in the middle third of the vocal tract (i.e., front-back tongue body position), and \(F3\) is lowered by very local, articulatorily peculiar constrictions at nodes of the \(5/4\) wavelength standing wave. Also, bilabial closure,
constricting the open-end antinode of each formant, lowers all formant frequencies; and front-velar closure, close to an F2 node and an F3 antinode, raises F2 and lowers F3 towards each other. We explained the heretofore mysterious, central fact about stop-vowel coarticulation, the so-called “locus” effect for coronals: a constriction, located between the open-end antinode and the palatal node of F2, shifts the sum of antinode and node constrictions towards, but not all the way to, a balance, so that relatively high F2 falls and relatively low F2 rises. Finally, in a small speech production experiment, we verified the theory’s prediction that the F2 locus for dentals should be lower in frequency than for more retracted coronals, since the location of constriction adds a greater component to the F2 antinode sum for dentals than more posteriorly articulated coronal stops.

In this derivation of formant structure from the shape of the resonating vocal tract, we have explained many of the most interesting, most complicated, most well-documented patterns in phonetics.

This quite old theory has interesting implications for current phonetic and phonological theory. For example, in important recent work on vowel articulation and phonology, Wood and his colleagues have challenged the traditional basic vowel dimensions, height and backness, arguing that the model of Stevens and House (1955), taken up also in Fant (1960), is both phonetically and phonologically superior. Their non-traditional account of vowel phonology, claims that “the vocal tract is narrowed at one of four locations: along the hard palate for [i-ε], and [y-ɑ]-like vowels, along the soft palate for [u-ʊ] and [u]-like vowels, in the upper pharynx for [o-ɔ] and [ɛ]-like vowels, and in the lower pharynx for [æ-ɑ]-like vowels.” They claim that “one cannot deduce articulation by translating F1 into ‘height’ and F2 into ‘backness’ ”(Pettersson & Wood, 1987). However, this is a rather good summary of what we have just shown can be done. They themselves refer repeatedly to the tuning of the shape of the vocal tract in relation to the sensitive nodes and antinodes of the standing waves. It was their mention of the idea, in fact, that led me to pursue the theory of nodes and antinodes.

We have seen that the effects of the shape of the vocal tract on formant frequencies match very closely with the traditional vowel features: openness and changes in F1 are

---

14See the phonological discussion on page 73.
15For example, see Figure 3, Pettersson & Wood (1987), which displays the nodes and antinodes of F1 through F4 in the vocal tract, adapted from Chiba & Kajiyama (1941).
directly related, and F2 is a measure of the combined effects of tongue-body frontness and lip aperture. Thus their complicated articulatory theory of place-of-articulation for vowels seems unnecessary from an acoustic point of view.

Such a theory also seems unnecessary from a phonological point of view. In an economical and accountable description of phonetic/phonological dimensions, there should be a consistent interpretation of the dimensions in both articulation and acoustics. Articulatory dimensions should map uniquely onto acoustic dimensions, and vice versa. For example, different degrees of articulatory openness should translate unambiguously onto some acoustic dimension (F1), and indeed they do. It would be undesirable if different points on the articulatory dimension should map onto the same point on the acoustic dimension, so that if there were 5 degrees of height, for example, the third and fifth had the same phonetic value. Changes in an articulatory dimension should be realized by similar corresponding changes in the acoustic dimension, no matter where on the scale. Doubling back is undesirable.

This criterion eliminates many possible articulatory-phonetic dimensions from consideration as phonological dimensions. The theory of node-antinode constrictions implies that constriction at a node is acoustically equivalent to widening at an antinode, though the articulatory configurations are different. It implies that “there is an essential many-to-one mapping from articulation to acoustics.”16 Only those articulatory dimensions which do have an unambiguous acoustic interpretation are acceptable, according to this criterion.

The theory that vowels have a place of articulation which varies from the palate back to the low pharynx has just this kind of undesirable property. As the place of constriction moves back in the vocal tract from the palate to the glottis, the acoustic effect on F2 is first to lower it (while moving the constriction towards the F2 antinode in the upper pharynx), and then to raise F2 (moving past the antinode, towards the node at the glottis). So the mapping from the articulatory dimension to the acoustic dimension actually changes direction: moving the constriction back can have opposite acoustic effects. Thus the place-of-articulation theory of vowels not only is overly complicated from the point of view of predicting acoustic structure, but also has an inconsistent mapping from articulatory to acoustic dimensions.

16 Thanks to David Graff for this formulation.
The theory presented here of the articulatory-acoustic mapping amounts to a reaffirmation of the traditional phonetic/phonological dimensions of openness and frontness (or height and backness; these are opposite sides of the same coin). However, some differences between the traditional articulatory dimensions and these dimensions which are simultaneously articulatory and acoustic should be noted.

“Openness” is a property of the acoustic tube formed by the vocal tract, not a tongue feature or a jaw feature alone, but a feature that derives from both tongue and jaw raising at the same time. The gradual widening of the tube at its open end along with gradual narrowing at its closed end are what constitute openness. It is not uniquely a lingual dimension nor a jaw dimension, but a dimension that measures their joint effect on the shape of the tube.

In several of the linguistically distinctive phonetic patterns discussed in this chapter, it was found that unrelated articulations had the same acoustic effect. The most striking confirmation of this was the predicted and observed pattern of three places of articulation for [ɔ]: labial, post-palatal, and low pharyngeal. Constrictions at each of these places have the same acoustic effect, that of lowering F3 by constricting at each of the velocity antinodes of the third standing wave in the vocal-tract.

Similarly, the raising of F3 sometimes found for [l] is consistent with a double apical/velar articulation of [l], where the constrictions are at points in the vocal tract where [ɔ] has the greatest widening: at nodes of the F3 standing wave.

Other cases support the same conclusion. For example, the usual co-occurrence of lip rounding with tongue-body backing has the joint effect of lowering F2, by constricting at both of the antinodes of the second standing wave. The articulatorily unrelated constrictions have the same acoustic effect. It almost seems that the articulation is designed to make the acoustic effect; it is put together out of separate pieces, as it were. It is no coincidence that the lips are often spread in the production of the extreme front vowel, [i], since widening of the tube at the F2 antinode that is located at the lips has the same acoustic effect as constricting the tube at the F2 node at the palate. Extreme tongue-body frontness has the same effect on F2 as lip-spreading. Again, separate articulations are joined for a similar effect.

Between Jakobson, Fant and Halle (1954) and Chomsky and Halle (1968), the phonetic
substance of phonological features was shifted in the theory from acoustics to articulation. This chapter might seem to suggest a great leap backwards, at least in the analysis of vowel sounds. Here we find that the simplest dimensions for vowels are acoustic, not articulatory, while quite complicated simultaneous articulations are found to occur whose best explanation seems to be the acoustic effect that they produce. However, rather than hopping back and forth over the articulatory/acoustic fence, let us instead simply recognize that the mapping from articulation to acoustics is causal, direct and simple. And if the patterns of constriction of the tube formed by the vocal tract are symmetric with respect to the nodes and antinodes of the standing waves in the tube, then the mapping is invertible as well.

Thus the measurements of the the first two formant frequencies, F1 and F2, that constitute the bulk of the phonetic data of this thesis, are to be taken as not just acoustic measurements, but phonetic measurements, which directly reflect articulatory dimensions of vocal-tract openness and tongue-body frontness/lip aperture. They are not mere acoustic epiphenomena, that are articulatorily uninterpretable. They are objective and rather precise measures of the fundamental, continuous phonetic dimensions of vowel quality, which are directly realized both in articulatory patterns and in acoustic patterns.
Chapter 3

Phonological Preliminaries: The Vowel Structure of Reference American

The goal of this chapter is to lay a linguistic foundation for studies of surface phonology and of the phonetics of speech performance in English dialects.

The chapter applies the theoretical machinery and principles of modern phonology to the analysis of surface phonological structure in English. The analyses of vowel structure of English dialects in this and other chapters develop interesting and surprising insights into the differences and phonological and phonetic changes relating English dialects. I have attempted to bring dialectal variation, surface phonology, and modern phonological theory together in a useful way.

Most recent studies of English vowel phonology are morphophonemic in nature; morphological alternations are the evidence used for underlying forms, and the structures argued for are considered to be part of the lexical phonological structure of the language. This chapter primarily relies on complementary distribution, phonetic similarity, and contrast for evidence about linguistic structure. The level of structure described is the relatively less-studied "surface" or post-lexical phonological structure (in the sense of Kiparsky 1982).

In studying the phonetics and phonology of English vowels, the first step is to define the
objects of study. First, the term "English" itself requires clarification; rather than studying the supposedly real but actually imaginary dialect of "General American English", I will describe an unabashedly imaginary dialect, "Reference American". They are quite similar, but the assumed, and false, reality of the former is not to be attributed in the same way to the latter. The phonological structure of Reference American, or RA, is summarized in the last section, and used as a basis of comparison with each of the dialects studied in later chapters.

The main question asked in this chapter is, What are the vowels of English? An enumerated list of phonemes or word-classes provides one kind of answer to this question. A more principled answer would describe the temporal and the static structure of English vowels, and locate vocalic classes in this structure. I argue for the specifically temporal character of phonological representations in Appendix 1, "Time in Linguistic Structure". First, what is their internal temporal structure? How many phonological segments may each vowel be divided into, and where do those segments fit into the structure of syllables and larger units? Once we have an idea of this temporal structure, we may go on to ask about the static structure: For each position in the syllable, what are the phonological features that distinguish among English vowels? How do these features combine?

After laying to rest the concept of "General American" and justifying Reference American, the chapter has three main parts: an enumeration of English vowels, an argument for a particular view of their temporal structure within a theory of syllable structure, and a discussion of the static features which are available to distinguish the classes at each location in the larger structure.

### 3.1 "General American English"

Linguistic ethnocentrism is pervasive. When scholars and lay people identify a "General American English" (GAE) they generally mean, "the way I talk", whether they are from Buffalo or Dallas or Philadelphia, just as speakers of modern Arabic dialects, when asked
where the purest Arabic is spoken, frequently give their own place of origin (Ferguson, p.c.). This dialect, which everyone seems to believe in, is in fact no real dialect at all. Does General American have a distinction between the sound classes in cot and caught?² Between morning and mourning? Between those in Mary, and marry? Or Mary and merry? Answers to these questions are not to be found, because General American, as popularly conceived, is not a well-defined object, either linguistically, socially, or geographically. Speakers of this supposed dialect may come from Texas or New York. But the dialects of Texas and New York are not phonetically identical.

The closest relation to reality of “standard” or “General American” English, is often to be found in the self-monitored, self-perceived speech of the individual speaker who is using the term. Explicating the term “standard” in this way is controversial, since most people who use the term think they are referring to some actual dialect. Phonologists, phoneticians, linguists and lay people all refer to this imaginary dialect. While it is certainly conceptually possible, its existence, much less linguistic or phonetic uniformity, has not been established.

An actual dialect is the speech form used by a speech community for communication in its day-to-day linguistic activities. It is not to be defined as the self-monitored imaginations of academics. The term, “General”, seems to imply that GAE is a uniform dialect spoken throughout the U.S. This is contrary to the facts. The work of dialect geographers and sociolinguists for the past 70 years has established, if nothing else, that standard, or General American English does not exist. Hans Kurath, after conducting a large-scale study of American dialect geography, stated: “The widely accepted assumption that there is a ‘General American’ type of English proves to be equally unfounded in fact; no Southerner or New Englander would ever have made such a generalization.” (1949:vi) He established Northern, Midland, and Southern as three major, distinct dialects of (Eastern) American English: “There is an extensive Midland speech area that lies between the traditionally recognized ‘Northern’ and ‘Southern’ areas.” (1949:v)(emphasis in original.) Depending on

²Apparently the answer is both yes and no. Peterson & Barney’s (1952) study of General American deserves mention in this context since the data from that study is continually being recirculated (e.g., Watrous 1991). They collapse together speakers of several dialects differing, for example, in whether they have or don’t have the phonological distinction between /á/ and /o/ as in cot and caught. The rather anomalous overlap between measurements of these classes (Figure 8) even in highly-monitored speech is therefore not very surprising.

38

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the differences that are considered to count as crucial, more or fewer dialects may be distinguished.

For some, “General American” could mean the conglomeration of all American dialects. But this step makes analysis of surface phonological structure impossible, since the different patterns contained in this conglomeration lead to very different structural analyses. Compare, for example, the structure of Alabama English given in that chapter with the structure of Chicago White English.

Research has shown that phonetic and phonological diversification of dialects is increasing, despite the widespread belief in the homogenizing effects of television, radio, and other modern forms of mass communication. Even relatively local subdivisions of these larger dialects, down to the level (at least) of large cities, may differ in important ways when phonetic/phonological behavior is studied. A merger may be in progress in one town, while twenty miles down the road, the phonological distinction is completely robust (Herold 1990).

Paying attention to geographic details is not yet enough; it still ignores the major social subdivisions within urban communities which are also phonetically highly differentiated. Within many U.S. cities, for example, the social division that is correlated with the greatest linguistic and phonetic differences is between (most) blacks and (most) whites. Where phonetic change is in progress, the speaker’s age may be systematically related to variations in phonetic details. Gender is another important variable, which has had more attention paid to it in laboratory studies than the variables of age and ethnicity. Finally, speech style is an important variable which has a significant influence on phonetic behavior. For example, speakers behave differently in important ways, often distorting their unmonitored (vernacular) sound pattern, when paying increasing amounts of attention to their speech (Labov 1966). Phonetic differences across these social dimensions have been found in vowel formant-frequencies (Labov, Yeager, and Steiner 1972), in consonant cluster articulation (Guy 1980), pitch range (Tarone 1973), etc. It is therefore impossible to analyse “General American” as a single linguistic object.

How can we best approach the study of American English dialects, then? One approach is to analyse each dialect on its own, from scratch. But no English-speaking linguist comes to the study of phonology without intimate knowledge at some level of a particular English
sound system. Another approach is to describe a single reference dialect in detail, and then show, for each other dialect examined, the differences between it and the reference dialect. Sound changes and sound correspondences may be insightfully explored by showing how the patterns of a particular dialect differ from those of a relatively well-understood reference pattern. This is true both at the level of phonological structure and of low-level phonetic realizations.

This comparison was implicitly done at the phonetic level, for example, in the foundational “Quantitative Study of Sound Change in Progress” (Labov, Yaeger and Steiner 1972). There, the differences between fine-grained phonetic realizations of sound-classes, found in the vernacular speech of particular speakers, was sometimes compared to the phonetic forms of a reference dialect. For example, if the nucleus of the /uw/ vowel in GOOSE overlaps in F1-F2 space with the nucleus of the /i/ vowel in KIT, it is said to be “fronted”. But “fronted” only has meaning relative to some reference point where the vowel is not fronted. In cases where older, more conservative speakers of the dialect were not studied, and unavailable for comparison, differences between the discovered patterns and those of an implicit reference dialect were taken as evidence of sound changes. Thus the reference dialect is implicitly considered to be the historical antecedent of the dialects studied. Such a claim is difficult to test directly, since we do not have much in the way of phonetic records for antecedents of these communities. Comparative reconstruction, to the extent that it has consequences for phonetics, may be able to shed light on the reality of such reference dialects, but until further work is done, we can only claim at best that this reference dialect is an inferred reference point that assists us in understanding how modern dialects got to be the way they are now. Perhaps further research will show that the inferred earlier dialects have more than inferred reality.

It is clearly very useful to be able to describe dialects in relation to such a reference dialect. However, “Standard English” or “General American” are names that unfortunately suggest a real, existing dialect — to repeat, a dialect that has been shown not to exist. Or if the many actual dialects that these terms correspond to are analysed as one, the result is an incoherent analysis. If some reference dialect is to be characterized, it should be a well-defined object, at least linguistically. This chapter phonologically defines a reference dialect of this kind, which I will call “Reference American”, or “RA”. This dialect is explicitly
a useful fiction. Like all phonologists that work on on their native language, my starting point is my own dialect. My Western U.S. dialect has a number of mergers between vowel classes that many English dialects keep separate. In order to characterize a system that is applicable to more dialects than my own, I include a number of distinctions in Reference American phonological structure which are absent in my own dialect. Some distinctions made in certain American dialects (for example, Southern *morning* vs. *mourning*) and in various non-American dialects are not included. The goal here is to describe a single, coherent, synchronic, phonological system, not to include, willy-nilly, all the differences in all known dialects, which may co-occur in no single dialect.

Thus Reference American is that dialect from which much of American English can be derived, containing all the distinctions I know of that are likely to have occurred in a single relatively conservative American English dialect. It may be considered as a conservative form of (Northern) American English, or on the other hand as merely a convenient fiction. It is not the non-existent uniform dialect which all cultivated Americans speak. Many of the differences among English dialects, which in this chapter are partly and temporarily ignored, are given their due in later chapters.

The point of this discussion is that we must be clear about the correspondence of what we study to actual, living, spoken dialects. Commonly they are two different things. Clearly, the goal of linguistic research is to elucidate the grammars of well-defined languages, rather than confused conglomerations of different dialects or unreal dialects which are claimed to be real. I do not deny the usefulness of convenient analytical fictions, since Reference American is such a fiction. But claiming that this dialect is one which is generally spoken by all (or educated, or upper-class, or Mid-Western, etc., etc.) Americans would be misleading.

We will find that the phonological structures of Reference American English are applicable to quite different dialects; the fact that the phonology of one dialect should be similar, in some fundamental ways to the phonology of another is quite intriguing. The differences between dialects are storable in terms of a small number of mergers and splits (often in particular environments), plus a number of differences in the rules which fill in the phonetic details of a rather underspecified phonological structure. In some cases, a

---

3I lack distinctive categories for /c, a:, ñerV/ as in THOUGHT, PALM, & marry.
considerable re-organization of the phonological structure is indicated (as in Alabama and Jamaican Creole, for example). This suggests the very strong claim that, despite the considerable differences in phoneme inventory and in phonetic realization rules, the principles governing the structure of the surface phonology of many English dialects are the same. This is supported by the observation that English speakers are often able to understand each other without a great deal of trouble throughout the English speaking world, despite their considerable differences.

Next I will enumerate a number of sound classes (and discuss the theoretical status of certain kinds of enumerations), and then describe Reference American surface-phonological structure. In later chapters I will apply the basic structures and dimensions argued for in this chapter to several quite different dialects.

3.2 English Lexical Sets

English vowels may be named and enumerated in many ways, depending on what one wishes to do with or to claim about the resulting list. One may enumerate "surface phonemes," "systematic phonemes," "word classes," "sound classes," etc. Among the central purposes of this thesis is to compare dialects, and for this purpose, the relevant conceptual entity is the "lexical set." John Wells (1982, vol. I) defined a number of lexical sets, which I will refer to frequently in this thesis. These consist of sets of words that are cognate in the different dialects; they provide an economical way of specifying the set of words which contain any given phoneme in any given dialect. The fact that these classes are listed and thoroughly characterized in a published work makes them especially useful, because this allows most words to be categorized unambiguously. Even more useful would be a dictionary which listed all words and the lexical set(s) that each belongs to.

Lexical sets are something like phonemes, and even more like the historical "word class" category used by Labov (e.g., LYS 1972:24ff). The terminology is initially disconcerting, since it implies that single sounds are identical with lists of words. All of the concepts, "phoneme", "lexical set", "word class", etc., may be understood as having a similar intensional and extensional meaning. The intension is "a linguistic class of sounds", and the extension is "the words containing that sound class". The names suggest one or the other
kind of meaning: "Lexical set" and "word class" suggest the extension, while "phoneme" or "sound class" suggest the intension. Nonetheless they all have, to this extent, the same meaning at both semantic levels.

The very important differences among them are the relevant domains of analysis: the set of words in the extension is claimed to behave as a group or unit in the domain of the concept: phonemes are a unit that (synchronic) phonological rules apply to; word classes are a unit that sound changes operate on; lexical sets are a unit that comparative statements refer to. Thus different approaches to linguistic sound classes make use of different terms.

Why make lists and sounds synonymous? A list of words is not a single sound! Even so, it is useful to refer to single sounds by the list of words which contain it, so as to avoid having to say exactly what the sound itself is. Suppose for example that a sound becomes something else across dialects or through history. Then a name (e.g., the IPA symbol) for the sound itself cannot be used to refer to this changing unit, since the name may be inappropriate once the sound has changed. In such a situation, as in the comparison of dialects or the study of historical change, the word-list approach to defining sound categories is more concrete and well-defined.

A historical word class is list of words which changes as a unit in the sound shifts, mergers, and splits operating on a dialect at a particular stage of history. If the dialect is traced farther back into history, to the time before certain mergers, say, then more word classes must be specified, since the classes that merged into one must now be distinguished with their own separate list. Thus the historical word classes necessary for analysing sound change in any particular dialect or language depends on the time period being studied and the classes which need to be distinguished in that period.

Similarly a "lexical set" may be defined as a list of words which constitutes the comparative categories of speech sounds that are necessary in the comparison of a particular set of dialects. More precisely, a lexical set is a list of cognate words for which, in all the dialects being described, all the words in the set contain a particular phoneme. Thus if the words containing phoneme A in one dialect have one of two phonemes B or C in another dialect, then there are two lexical sets, which might be labelled AB and AC. The first set contains the cognate words which have A and B in the two respective dialects, and the
second contains the cognates with A and C, respectively.

The purpose of dialect comparison is better served by comparative classes like Wells’ lexical sets than historical classes such as the Middle English inventory of vowel phonemes, etc., even though the totality of historical word classes which are relevant to the description of sound changes in any English dialect are never entirely identical to Wells’ enumeration of lexical sets. Without detailed study of the history of English phonology, references to historical word classes are largely incomprehensible. I do not assume that the reader is an expert in English historical phonology. The task in dialect comparison is to find out what lists of words correspond to the phonological classes that are distinctive in a particular dialect, and this is the task that lexical sets do.

For further discussion and justification of the concept of lexical sets, see Wells (1982, Vol I, p117ff). This group of lexical sets is somewhat arbitrarily chosen (it is the sets formed by the comparison of British Received Pronunciation and what Wells takes as “General American”, namely a conglomeration of different American dialects), but it is nonetheless quite widely useful. It is important to note that Wells does not include many lexical sets which are necessary for the comparative description of particular dialects, such as the sets corresponding to the meet/meat distinction in Hiberno (Irish) English (J. Harris 1985), the can (noun)/can (aux. verb) distinctions in New York City, and Philadelphia, and other differences, both ancient and new, in many dialects. New sets must occasionally be defined for comparison of particular dialects.

In Table 3.1 I list Wells’ lexical sets by name, in capital letters. Wells gives a detailed enumeration by environment of examples of each of his lexical sets. Examples of the classes that I find difficult are discussed below and given in Table 3.2, in hopes that this will aid the reader. The names for the sets were presumably chosen to be members whose pronunciation is not identical with the name of any other set in any dialect. Because of the many vocalic mergers in various dialects, the names include different consonants and on-glides, which serve the function of keeping them apart. Thus there should be no dialect in which the names for any two sets are homophonous.4

Some lexical sets do not correspond to single phonemes in any dialect. For example, in

---

4 The layout in this listing is designed so that sets that are not distinguished in one dialect or another are relatively close to one another, where possible. The sequence of columns may be described as: front-r-glides, front-glides, other front vowels, other vowels, back-glides, other-r-glides.
Table 3.1: Wells’ Lexical Sets

<table>
<thead>
<tr>
<th>NEAR</th>
<th>FLEECE</th>
<th>KIT</th>
<th>FOOT</th>
<th>GOOSE</th>
<th>CURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQUARE</td>
<td>FACE</td>
<td>DRESS</td>
<td>STRUT</td>
<td>GOAT</td>
<td>FORCE</td>
</tr>
<tr>
<td>PRICE</td>
<td>TRAP</td>
<td>LOT</td>
<td>MOUTH</td>
<td>NORTH</td>
<td></td>
</tr>
<tr>
<td>CHOICE</td>
<td>BATH</td>
<td>CLOTH</td>
<td>START</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PALM</td>
<td>THOUGHT</td>
<td>NURSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAPPY</td>
<td>COMMA</td>
<td>LETTER</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

British Received Pronunciation the vowel in Wells’ CLOTH set is pronounced the same way as the vowel in the LOT set, while most Eastern American dialects pronounce the CLOTH set with the vowel used in the THOUGHT set. Wells makes no claims of phonological reality for such sets; this does not reduce their usefulness both in this research and in exposition.

With any or no training, you can immediately identify the sound class in your own dialect which corresponds to the sound class in the dialect under discussion, simply by noting which of your phonological classes occurs in the word that names the lexical set. Of course, you cannot infer that all words with that phoneme in your own dialect are members of the lexical set referred to, since mergers and splits have rearranged your system as well as the one under discussion. When I see the name of the set, LOT, I know it refers to some of the words in my /a/ phonological class, though not all, since CLOTH words and THOUGHT words are also in that class of mine.

At least the first problem of reading comprehension, namely that of identifying the relevant phonological class in the reader’s own dialect, is solved better by using a word which contains the sound than a symbol which may refer to any of several possible classes in Middle English or some other dialect, which may be unknown to the reader. In practice, the differences between description in terms of historical classes and in terms of lexical sets are often quite small.

Wells’ lexical sets are used here to compare English dialects, the purpose they were designed for. Many of the phonological differences among English dialects may be described by stating which lexical sets are identified as the same in one dialect or another. For example, we may state a number of conditioned mergers and near-mergers\(^5\) which have

\(^5\)I discuss “near merger” below on page 306; also cf. Labov, Yaeger, & Steiner (1972, Chapter 6 and
occurred in particular American dialects in terms of Wells' lexical sets:

<table>
<thead>
<tr>
<th>Dialect</th>
<th>minimal pair</th>
<th>Lexical Set characterization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern</td>
<td>pen/pin</td>
<td>DRESS and KIT merge before nasal</td>
</tr>
<tr>
<td>Californian</td>
<td>Mary/marry/merry</td>
<td>SQUARE merges with TRAP, DRESS before /r/.</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>beer/bear</td>
<td>NEAR and SQUARE merge.</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>merry/Murray</td>
<td>DRESS, STRUT near-merge before /r/+Vowel.</td>
</tr>
</tbody>
</table>

The differences among lexical sets which I find most confusing are NORTII vs. FORCE, and CLOTHII vs. THOUGHIIT vs. LOT vs. PALM. NORTII and FORCE constitute a single sound class for me, but many dialects differentiate them (including Alabama and Jamaica, studied here). Similarly, CLOTHII, THOUGHIIT, LOT, and PALM all contain the same vowel in my dialect, but they are distinct for many. There are traditional names for three of these classes: THOUGHIIT is called “long open O”, LOT is called “short O”, and PALM is called “broad A”. CLOTHII is that set of words which is pronounced like the LOT set in British Received Pronunciation, and like the THOUGHIIT set in Wells’ “General American”. These sets are exemplified in Table 3.2.

Table 3.2: Examples of some confusing lexical sets.

| NORTII | morning, war, or, forty, horse. |
| FORCE  | mourning, wore, ore, four, hoarse. |
| PALM   | father, balm, spa, drama, schwa, rajah. |
| LOT    | bother, bomb, cot, doll. |
| THOUGHIIT | daughter, bought, caught, fall, saw. |
| CLOTHII| off, gone, lost, moth, long, coffee. |

Wells’ list of lexical sets provides, by stipulation, one answer to the question, “What are the English vowels?” This answer is justifiable for the task of comparing dialects, which is usefully approached using categories like these. But we may also ask for a synchronic, phonological explication of the concept, “English vowel”.

---

6 Of course, stating a merger using lexical sets does not show what the resulting phonetic forms are.

Appendix A).
3.3 Temporal Vowel Structure in English

What is the temporal, or sequential, structure of Reference American English vowels and vowel sequences? On the surface level, RA English has various kinds of temporally complex, or dynamic, vowels. It has long vowels (as in *eat* [i\(^\text{yt}\)]) and short vowels (as in *it* [i\(^\text{t}\)]). It has seven kinds of glides: on-glides (*wad* [wa\(^\text{d}\)]) and off-glides (*Dow* [d\(\text{a\textsuperscript{e}}\)]); up-glides and in-glides (both exemplified in *idea* [a\(\text{c\textsuperscript{d}\text{ia}}\)], back-glides and front-glides, and even r-glides, which go out into another dimension entirely. It has an indefinite number of “phthongs”: monophthongs (*it*), diphthongs (*out* [a\(\text{u\textsuperscript{t}}\)]), triphthongs (*our* [a\(\text{u\textsuperscript{r}}\)]), and n-phthongs (*wire* [wa\(\text{e\textsuperscript{r}}\)], *wiry* [wa\(\text{e\textsuperscript{r}\text{i}}\)], *wirier* [wa\(\text{e\textsuperscript{r}\text{r}\text{i}}\)], *you are wirier* [yu\(\text{a\textsuperscript{r}}\text{wa\textsuperscript{e\textsuperscript{r}\text{r}\text{i}}\text{r}}\)]...). In this section we reduce all these dynamic vowels or vowel-sequences to a simple underlying phonological structure.

We may begin this task by first considering a basic phonological unit: the syllable. Minimal pairs like *layer-lair, mower-more, see-see*\(^\text{7}\) may be phonologically distinguished as two syllables versus one.\(^\text{8}\) This fundamental category allows us to divide many classes of n-phthongs into sequences of syllables, each one of which contains vocalic segments from a much smaller set of n-phthongs, where n ≤ 3. If we exclude the supra-syllabic phenomena of stress, rhythm, etc., then the temporal structure of an n-phthong simplifies to the bare sequence of syllables which makes it up, and our question can be focused more sharply: “What is the temporal structure of RA English vowels within the syllable?” As in this first step, so later: temporal phonological structure is simply the unidirectional sequencing of abstract units.

This simplified view of time does not apply to phonetics. In phonetics, time is continuous rather than discrete, as it is in phonology — this is discussed at length in Appendix 1. Thus the temporal patterning of phonetic events is on a continuous scale. Temporal patterning is of two kinds: duration and “complexity”. Duration concerns how long things stay the same across time, and complexity concerns the patterns of change across time.

Duration has an absolute limit of zero; the shortest possible class of vowels are those that reach or approach the absolute limit of zero duration, or complete deletion. In fact,

\(^{7}\text{My monitored pronunciations of these are [le\(\text{c\textsuperscript{r}}\), le\(\text{c\textsuperscript{p}}\); mo\(\text{u\textsuperscript{r}}\), mo\(\text{u\textsuperscript{p}}\), si\(\text{r\textsuperscript{r}}\), si\(\text{r\textsuperscript{p}}\].}

\(^{8}\text{Bailey (1985:162) makes the same point.}

47

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
there is a significant proportion of vowels that are realized by so little phonetic substance that there is nothing to be measured (for example, 7.6% of 4821 vowels for a Chicago speaker). There are also a number of vowels whose acoustic manifestation is very, very short — as little as a single pitch pulse. There is in fact a continuum of duration reduction for vowels, governed by several factors including stress, position in the utterance, vowel height, and adjacent phonological environments (for recent studies, see Crystal & House 1990; van Santen & Olive 1990). Absence of stress correlates with the shortening of vowels. Utterance-initial syllables may be nearly or entirely deleted. Pre-boundary syllables are lengthened according to the strength of the boundary, so that even unstressed vowels may be quite long before pause. The more open the mouth, the longer the vowel. Between voiceless consonants, vowels often devoice: no acoustic vowel (that is, a voiced vocalic acoustic segment) may remain. Other effects on duration or length than these may also exist. These factors are evident in results, presented in Appendix 3, derived from acoustic vowel duration measurements done for each of the speakers examined in this thesis.

As for complexity, it is clear that such short vowels have little time to be complex: measurable gliding can hardly occur within a single pitch pulse. When the short vowels discussed below occur in lengthening environments in some U.S. dialects (stereotypically in Southern dialects), such as in stressed monosyllables, they develop phonetic inglides, so that they are phonetically more complex than the same vowel classes in non-lengthening environments.

3.3.1 Degrees of Phonological Vowel Length

Restricting the discussion now to phonological temporal structure, I will examine evidence for three degrees of vowel length and complexity, as opposed to the traditional two degrees of long vs. short. In the end, I will reject the claim that there are three degrees of phonological vowel length in English, but not before the evidence relevant to this question leads to a fundamental re-thinking of the structure of the English vowel system.

---

9 Thus a secretary recently heard answering the telephone as “Peter Science” was reinterpreted correctly as saying “Computer Science” after normalizing for utterance-initial syllable deletion, /aw/-fronting, and the physical location of the speech event. Or for example in some speech recorded in Southampton, England, an utterance-initial syllable was constituted by a single pitch pulse, 40ms before the onset of the audible part of the utterance. Such observations may be typical of naturally spoken conversation.
The first degree of phonological vowel length or complexity is exemplified by the monophthongal\textsuperscript{10} "short" vowels, as in *pit, pet, putt, put*.\textsuperscript{11} These vowels do not occur in open stressed syllables: when lexically stressed, they must be followed by a tautosyllabic consonant.

The second more complex (or in this case, longer) set of vowels are the traditional mid and high "tense" vowels, which occur in both free and checked syllables, as in *be, bean; bay, bane; Boo, boon, Bo, bone*.\textsuperscript{12} Giving acceptable names to this class of vowels and to the difference between it and the first class of vowels above is a problem that has occupied phonologists and phoneticians for decades. They were traditionally analysed as "long" vowels, as opposed to the "short" vowels above. This discussion will use these terms, though each of the other terms has some value. They have been called "tense", as opposed to "lax", a term which is applied to the short vowels. The tense/lax distinction is notoriously ill-defined. Bloomfield (1934:101), for example, says the "simple [i.e., short] vowels are pronounced with loose muscles...[while] [ij, ej, uw, ow] seem to be somewhat tense throughout", but overall vocal-tract muscle tension does not necessarily differentiate these vowels. Labov abandoned the term, "tense", and proposed the acoustical term, "peripheral" (as in "peripheral in acoustic vowel space") to replace some of its senses. Since there are also non-peripheral, long vowels, such as stressed /æ/, raised /ay/, etc., which are pronounced centrally when articulated with the greatest force, and which count as long vowels, I will not use this as an underlying phonological dimension. Bailey (1985:205) calls long vowels and diphthongs "heavy", and short vowels "light". Mora-counting also differentiates them as one versus more than one morae. Wells uses the distributional terms "checked" (never syllable-final and stressed) vs. "free" (sometimes syllable-final and stressed) to differentiate them. This is the main distributional fact that holds most English dialects together as similar to one another.

The vowels I call "long" vowels may be considered "longer" than the "short" vowels according to the following metaphor. Suppose that, in order to be stressable, syllables require a certain quantity or length of segmental material. Then, vowels that occur only in

\textsuperscript{10}As mentioned above, these sound classes often have phonetic inglides in the Southern U.S. — cf. the Alabama chapter — but not in the basically Northern dialect of Reference American.

\textsuperscript{11}Phonetically, these may be written [pʰɪt], [pʰɛt], [pʰʌt], [pʰʊt], respectively.

\textsuperscript{12}[bi], [bɪn]; [be], [be ʰn]; [bu ʰ], [bu ʰn]; [bo ʰ], [bo ʰn], respectively.
closed syllables when stressed may be considered “not long enough” for this purpose, since additional following segments (which close the syllable) are required to form a stressable syllable. Vowels that can occur in open syllables on the other hand are “long enough” in this sense, since they provide the necessary quantity of material to form a stressable syllable on their own. This explicates the traditional distinction between “short” and “long” vowels, a distinction which is based on this distributional difference. The phonological picture is paralleled by the phonetic facts: these “tense” and peripheral vowels are indeed longer in duration, other things being equal, than their “lax”, non-peripheral counterparts. Further justification for the term “long” is that these vowels are indeed phonetically long monophthongs in various dialects.

A third set of vowels, however, is even “longer” than the other classes, according to a similar metaphor. This class consists of the diphthongs in *buy, boy, bough*. This class is not distinguished from the “long” vowels by the checked vs. free distribution, since both occur in checked as well as free syllables (e.g., *heed, hide, he, high*). They are different from the mid- and high- long vowels in that they occur in a smaller set of (checked) environments. In particular, postvocalic /ɾ/ cannot follow them within the syllable. Consider the distribution of post-vocalic /ɾ/ with the merely “long” vowels in Table 3.3.

Table 3.3: Contrasts between one versus two syllables.

<table>
<thead>
<tr>
<th></th>
<th>rear</th>
<th>free-er (more free)</th>
<th>tour</th>
<th>doer</th>
<th>lair</th>
<th>layer</th>
<th>more</th>
<th>mower</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The contrast between the two columns, as mentioned earlier, may be attributed to syllable count: one versus two. But when I try to find such pairs for the third class of vowels, I fail. For all the speakers whom I have asked (of those whose dialects are r-ful

---

13See also Peterson and Lehiste (1960) on “intrinsic vowel length”.
14Los Angeles Chicano, Jamaica Creole, and Wisconsin/Minnesota, to cite three dialects, often have monophthongs for the “long” vowels. Cf. also the duration measurements in Appendix 3.
15[ba水源], [bo水源], [ba水源], respectively.
16These may be written phonetically as [hi AudioSource/d], [hi AudioSource/d], [hi AudioSource/e], respectively.
17These may be transcribed as [i AudioSource/e], [i AudioSource/e], [mo AudioSource/e], [mo AudioSource/e], [i AudioSource/e], [i AudioSource/e], [i AudioSource/e], [i AudioSource/e], [i AudioSource/e], [i AudioSource/e].
and have a front-raising glide in *buy*), the pairs *hire* and *higher*, *mine* and *(Oscar) Mayer* are identical.\(^{18}\) A root-beer company promoting its connection with the Philadelphia ice-hockey team prints on its cans the phrase, “Hires Flyers”, which rhyme suggests the same identity.

While phonetic evidence must be interpreted cautiously, the following spectrogram (Figure 3.1) of the phrase *fires 'n' wars* (taken from some of the vernacular speech analysed in the Chicago chapter) shows two relevant examples. These are in fully stressed, list intonation, in the longer context, *It's lasted through, fires, and wars, and God knows what.* (where “,” signifies final lengthening and a fall in pitch), so both /aɪr/ and /ɔr/ occur in (final) lengthening environments. An impressionistic transcription of this sequence is \([fə'zəːraʊz\ θ]\). The first word, *fires* contains an extremely long, steady-state [ə] segment (\(\sim 215\)ms, measuring from the midpoints of the transitions from and to adjacent segments), while the second contains an /r/ realization whose formant structure changes continuously and which is only \(\sim 140\)ms in duration. These phonetic differences are consistent with phonologically syllabic and non-syllabic /r/ for *fires* and *wars*, respectively.

Similarly, *hour* rhymes with *power*, and *flour* and *flower* are identical.\(^{19}\) Even *our*, a closed-class word which may be expected to reduce as much as possible, either rhymes with *flower*, or else it is identical with *are*, where the high-back glide is lost along with the syllabicity of the /r/. I find no examples of monosyllabic /aɪr/ sequences which might correspond to *foyer* and (in my dialect) *lawyer*, etc.

Since /r/ cannot follow these diphthongs within a single syllable, we may infer that the position in the syllable that post-vocalic /r/ occupies is filled up. Thus these diphthongs fill up more of the syllable than the mid- and high- long vowels in *see, say, sew, Sue*\(^{20}\) which do allow following /r/ in the same syllable (as in *beer, bear, bore, boor*\(^{21}\)). This class is thus different from, and longer than the “long” vowels. Our metaphor may now therefore be re-examined: The three classes, “short”, “long”, and “longer” are, respectively: not long enough to close a stressed syllable, long enough, and so long that some segments can’t fit

---

\(^{18}\) For me, these all end in \([æ^\varepsilon\ θ]\).

\(^{19}\) *Flour* and *flower* may be a poor example, since they are historically derived from the same word, as in “the flower of the wheat.”

\(^{20}\) *[siɛ], [se^\varepsilon], [s\ θ], [su^\varepsilon]*, respectively.

\(^{21}\) *[bi^\varepsilon], be^\varepsilon, bo^\varepsilon, bu^\varepsilon]*.
Figure 3.1: Judy from Chicago: "...fires, 'n' wars ..."

in the syllable with them. Table 3.4 summarizes the constraints.

Table 3.4: Constraints on consonants after stressed vowels

<table>
<thead>
<tr>
<th>Vowel is</th>
<th>Post-vocalic consonants are required</th>
<th>Post-vocalic consonants are restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>short</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>long</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>longer</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

These three degrees of apparent vowel length will be explicated in a theory of the temporal structure of vowels within the syllable in the next sections.

Yet another degree of length might be added by considering vocalic on-glides such as /y, w, r/. However, this leads to uninteresting questions like "Is the vocalic part of yet (which could be written /ye/) longer or shorter than the vocalic part of ate (which could be written /ey/)?" Also on-glides are more consonantal than off-glides.\footnote{For example, as pointed out in Chomsky & Halle (1968), and Janda (1988), German on-glides /w,y/} We will not
discuss on-glides further. These facts would appear to support the idea that there are three phonological degrees of vowel length. The basic linguistic observation is that there is a complementary distribution, within the syllable, of the glides in buy, bough, boy and postvocalic /r/. This observation may be interpreted in a different way.

3.3.2 Length and Complexity in Structuralist Treatments of English Vowels

The complementary distribution pointed out above suggests that there may be a position of some kind in the syllable for /r/ (and for the glides in the diphthongs of buy, bough, boy). Let us explore this idea, by examining a theory of English vowels which takes this position quite seriously. We will first present the theory, and then point out certain internal problems and some modifications that are necessary in order to make the theory compatible with the above-given observations about vowel length and complexity and with current views of syllable structure.

Since The Sound Pattern of English (henceforth SPE), treatments of English vowels as sequences have been out of fashion: the diphthongs in high, and how, for example, were derived from monophthongs by rules that recapitulate the Great Vowel Shift. But before SPE, English vowels were defined in terms of sequential structures involving slots for particular classes of segments. Characterizations of English vowels in terms of temporal structures (i.e., vowel-vowel sequences) may have incorrectly been rejected because of the emphasis on static phonology. With the return of the syllable to phonological theory, we may take up this work again.

Consider the American structuralist description of (American) English vowel systems. This trend reached a pinnacle in the work of the George Trager with Bernard Bloch (1941), and with Henry Smith, in (1951).\footnote{But see also, \textit{inter alia}, Bloomfield (1934), Whorf (1943), Swadesh (1947), Pike (1947).} Take, for example, the treatment of the structure of English syllabic phonemes by Trager & Bloch (henceforth T&B). They establish 6 “short” vowel phonemes, called /i, e, a, o, ø, u/. These short vowels are concatenated with other phonemes to form temporally complex sequences of the following 8 classes or “sub-systems”: /V/, /Vj/, /Vw/, /Vh/, /Vr/, /Vjr/, /Vwr/, /Vhr/. V represents one of the

\begin{itemize}
\item become fricatives, but off-glides do not. Thus /au/, /al/ don't become [av], [aj].
\end{itemize}
short vowels; /j,w/ are up-glides to the front and back, respectively (I will use /y/ for /j/, below); /h/ is a centralizing off-glide, which can also represent length; /r/ is the rhotic semi-vowel or its non-rhotic reflex in "r-less" dialects. The term "sub-system" will often be used to refer as a unit to sets of vowels with the same glide. As LYS defines it:

“By sub-system, we mean a set of vowel nuclei or peaks which have the same glides or satellites and the same super-segmental features. Vowels within a sub-system differ only in the three dimensions of [height], backing or rounding.” (p. 219)

Labov (1989) shows that a greater degree of confusion or misunderstanding occurs in natural speech comprehension between vowels within sub-systems than is found between vowels across sub-systems.

Many of the possible combinations of the short vowels with the glides, constituting different sub-systems, are exemplified by large classes of words; others are exemplified by small word classes with as few members as one; yet others are entirely empty. The following table is taken from T&B (p. 243), substituting the names of Wells’ lexical sets in capital letters for T&B’s example words in those cases where the vowel classes correspond.24

Table 3.5: Trager & Bloch’s English Vowel Structure (1941)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>/V/</td>
<td>/Vw/</td>
<td>/Vh/</td>
<td>/Yr/</td>
<td>/Yjr/</td>
<td>/Vwr/</td>
<td>/Vhr/</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>KIT</td>
<td>FLEECE</td>
<td>idea</td>
<td>mirror</td>
<td>NEAR</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>DRESS</td>
<td>FACE</td>
<td>yeah</td>
<td>merry</td>
<td>eyrie</td>
<td>—</td>
<td>SQUARE</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>TRAP</td>
<td>PRICE</td>
<td>MOUTH</td>
<td>PALM</td>
<td>marry</td>
<td>Irish</td>
<td>cowrie</td>
<td>START</td>
</tr>
<tr>
<td>o</td>
<td>LOT</td>
<td>CHOICE</td>
<td>THOUGHT</td>
<td>sorry</td>
<td>Moira</td>
<td>—</td>
<td>FORCE</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>STRUT</td>
<td>GOAT</td>
<td>huh</td>
<td>hurry</td>
<td>—</td>
<td>—</td>
<td>NURSE</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>FOOT</td>
<td>GOOSE</td>
<td>—</td>
<td>jury</td>
<td>—</td>
<td>CURE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These combinations may be more economically expressed as in (1), or more abstractly as in (2):

24The lexical sets with no place in this table are six: CLOTH, which goes with either THOUGHT or LOT, depending on dialect; BATH, which goes with either TRAP or PALM, depending on dialect; NORTH, which goes with FORCE in most dialects; and HAPPY, LETTER, COMMA, which are lexically unstressed and therefore excluded from T&B’s consideration.
Labels for the first two positions are somewhat arbitrary. What is essential to T&B's claim is that these positions are distinct and can be filled only by the corresponding list of segments. I have chosen to label them with familiar phonetic terms, "Nucleus" and "Glide". "Peak" and "satellite", or other terms might also be used.

The distinctions made in T&B's analysis are insufficient to represent the contrast discussed above between long vowels /iy, ey, aw, uw/ and longer vowels /ay, aw, oy/. For example, there is no formal difference of length between /ey/ and /ay/. If /r/ occurs in a position following the glide slot within the syllable (that is, in the syllable coda), why shouldn't it occur there in words like hire, flour? It seems that it cannot, since these words break into two syllables, merging with higher, flower. Further, there are words like heist, mounds, etc., which do have segments in the coda. If /r/ is a coda consonant, and coda consonants are acceptable after the longer vowels /ay, aw, oy/, then monosyllabic forms like /ayr/, /awr/, /oyr/ should be possible. However, they are not.

What, then, is the difference between long vowels /iy, ey, uw, ow/ and longer vowels /ay, aw, oy/ that accounts for this contrast? One difference is that the latter are low vowels in some dialects. However, the true answer to this question will become clear only after we clarify the analysis of /r/ within the syllable, as is done in the following section.

### 3.3.3 Is /r/ a Glide?

Trager & Bloch

I will argue that the glide slot need not be kept separate from the slot for /r/. Let us reconsider Trager and Bloch's chart, Table 3.5 above. The last three columns of this chart constitute a catalog of counterexamples to the claim that /r/ occurs in glide position, since they each contain a glide preceding /r/. Three observations will suffice to eliminate these counterexamples.

---

25 These symbols are constructed from T&B's table. /aw/ is often also written as /ow/.

26 However, /oy/ is raised to mid or higher in many dialects including my own; I therefore analyze it, below, in Reference American, as containing a mid-back nucleus.

---
First, note that the /r/ in the disyllables eyrie, Irish, Moira, and cowrie may be analysed as syllabified with the following syllable. Indeed if the observations about /ay, aw, oy/ above is taken seriously, this step is required, since /r/ cannot occur in the same syllable with just these vowels. This step eliminates several phoneme-classes.

Second, notice that once these words are excluded, Columns F and G (Vyr, Vwr) have only a single member each, and perhaps coincidentally both of them correspond to gaps in column H (Vhr). If this complementary distribution is not a coincidence, then columns F and G should be collapsed with column H, resulting in a pleasantly gap-less column. Gaps in a paradigm are often suggestive of missing generalizations; this generalization eliminates twelve.

Thus the separation of the ({y, w, h}) slot from the (r) slot has been deprived of two-thirds of its justification. If the Vhr class of vowels in column H can somehow be removed, then T&B’s justification for the separation of glides from /r/ will be completely destroyed. This leads to the third observation.

Notice that the Vr class of vowels all occur before unstressed syllables, where the /r/ may be ambisyllabic. None of the Vhr class occurs in this environment; the /r/ in this class is tautosyllabic. This is yet another case of complementary distribution of classes within the structure; it allows us to collapse the two classes into one. The fact that the Vhr class is phonetically more tense or peripheral than the Vr class must then be accounted for by a phonetic rule which tenses these vowels before /r/ in the same syllable. The rule is familiar from previous studies of English vowels:27 the independently motivated vowel-before-vowel raising rule (which converts the vowel following /d/ in idea into a peripheral vowel, for example).

Thus we have been able to eliminate all 18 of T&B’s catalog of counterexamples to the theory that /r/ occurs in glide position.

In a later work, Trager & Smith (1951) attempt to provide a pan-English vowel structure. By this time their greater knowledge of phonetic differences among English dialects, combined with the American structuralist principle of bi-uniqueness, resulte4 in an even greater proliferation of unnecessary phonemes. The 1951 system contains three heights and

---

27For an early statement, Kenyon and Knott's section, Variations, §92, p. xxxviii, in the Pronouncing Dictionary which has served as the data for much of the generative phonological treatment of English. For recent discussion, cf. Halle & Mohanan 1985, Section 2.
three front/back levels, making nine base vowels, each occurring in eight sub-systems (V, Vy, Vw, Vh, VrV, VrC, Vhr, VhrV). Because the analysis is pan-English and bi-unique, vowels with following /r/ in "r-ful" dialects are analysed as occurring in different locations in the structure as phonetically different "r-less" phonemes, thus accommodating the "r-less" dialects at the same time, in the same system. Nine phonemes in their analysis are vowels before /r/ in dialects where /r/ is lost — but they are considered to be in the same system, rather than in the new and separate systems of the r-less dialects. A result of this approach is that the basic correspondence between sound class and lexical set or word class is lost: some word classes correspond at the same time to multiple sound classes, since in different dialects they are pronounced in different ways. Further, a number of evidently allophonic differences are analysed as phonemic, as between jury and boorish, story and pouring (where a preceding labial and following word boundary may for some speakers make the vowel in the boorish and pouring phonetically longer and backer than the vowel in the other contexts). Their VrC and Vhr sub-systems contain the identical word-classes. Though their considerable knowledge of English dialects and their sensitivity to subtle differences of sound makes Trager and Smith (1951) a very useful source of information about sound patterns in English, their theoretical assumptions lead them away from, rather than towards, the kind of system argued for in this chapter.

Yes, /r/ Is a Glide.

In the interests of clarity, let us consider the nucleus, glide, and coda to be mutually exclusive, so that the coda contains the post-vocalic segments other than /y,w,r/. In Section 3.3.4 below I discuss the issue of how these three elements fit into syllable structure. For now, let us consider them as non-overlapping, distinct constituents, which occur at most once in each syllable. This will make it easy to map them onto the correct syllable structure representation later.

In the schema derived from Trager & Bloch (1941), namely Nucleus (Glide) /r/, the vowels, /i, e, a, o, a, u/, are in the Nucleus, /y, w, h/ are in the Glide, and /r/ forms a slot of its own, distinct from and freely combining with the others. Let us suppose instead that:
(3) /r/ occurs in the Glide position.

(4) Glide position supports at most one segment.

Are (3), (4) compatible with the observations made above of the three degrees of phonological vowel length? The answer is no; the facts from which the third degree of length was derived will be explained by (3), (4), but the view in which Reference American vowels have 3 phonological degrees of length must be abandoned, as I now show.

The diphthongs /ay, oy, aw/ unambiguously contain a glide.28 It then follows from (3), (4) that /ay, oy, aw/ cannot be followed by an /r/ within the same syllable. The glide position is already filled, so by (4), /r/ cannot occupy it. By (3), /r/ cannot occur in coda or nucleus positions either. So if /r/ is to be licensed by syllable structure, it must go in a separate syllable. (It may also be deleted, as in r-less dialects where the vowel in hire is a monophthong, as [ha:]. Then, it can also be a monosyllable, which is — not coincidentally — compatible with the theory’s constraints.) As observed above, breaking occurs in Reference American in words with glide-r sequences (dire, hour, etc.) so that the form becomes disyllabic. Thus the theory, which states that only one glide can occur in a syllable, and that /r/ and /w, y/ are glides, is consistent with the facts.

A final observation re-confirms the analysis of /r/ as a glide, distinct from the coda. Various coda consonants, even consonant clusters, do occur after /ay, aw, oy/ within the syllable (as in heist /hayst/, mound /mawnd/, joint /joiynt/, etc.), though /r/ cannot. If /r/’s location is an undistinguished part of the coda, which evidently may contain final nasals and obstruents, then /r/ should also occur in the same environments. But /r/ is not acceptable in these environments. An additional stipulation is necessary to restrict /r/ from these contexts while allowing other coda segments to occur there. This stipulation is avoided if /r/ is analysed instead as a glide.

What remains of the short/long/longer distinction discussed above? Vowel length is here expressed as sequential structure. Thus the class of longer vowels, /ay, aw, oy/, are formally represented expressed as nucleus-glide sequences. The class of short vowels, as in pit, pet, Pat, pot, putt, put,29 are bare nuclei, without any glide. What about the

28In many Southern U.S. dialects in various contexts, /ay/ has monophthongized, so the high-front glide is absent. In Reference American, this glide is present.
29These may be pronounced as [pUt, p£t, pæt, pUt, pæt, pUt].

58
long vowels? These are also expressed as nucleus-glide sequences when T&B write them as /iy, ey, ow, uw/, making no formal distinction of length between the long and longer classes of vowels. The next observation will show, first, how the long/longer distinction is unnecessary to account for the different behavior of the vowels before /r/, and second, how the independently motivated analysis of /r/ as a glide can be used to clarify new facts that are otherwise inexplicable.

The facts are best introduced in their historical context. In England around the beginning of the 17th century, the vowels in the words her, fir, fur, and unstressed for\textsuperscript{30} merged into a single, continuant, rhotic monophthong\textsuperscript{31} without distinctive [i, u, e, o] color — that is, without distinctive height or backness. (The merger actually took place in two steps, according to Jesperson 1909:319. First the high vowels merged: fir = fur; then the mid vowel: berth = birth.) Thus the former length contrast among non-low vowels before /r/ (exemplified as fear vs. fir, hair vs. her, boor vs. burr) was lost, and is not now distinctive. The present contrast between these vowels is in whether or not they have nuclear /ə/.

A phonetic explanation for this response to the change may suffice: the retroflex continuant /r/ gesture “has a relatively incompressible trajectory whose timing ‘slides’ with respect to other gestures in a sequence” (Boyce & Espy-Wilson 1991). The effect of this “sliding” is that the /r/ gesture may begin even before preceding segments begin, so as to attain a configuration that produces a fully rhotic sound at the point in the acoustic sequence corresponding to the timing slot of the /r/ itself. The temporal incompressibility of this very complex gesture (which can involve three constrictions, as discussed on page 29 and by Ohala 1985) suggests that the rhotic quality may bleed into the preceding short vowels, merging them together, while preceding long vowels were long enough to retain their individual nuclear vowel qualities.

Today the situation remains the same: There is no contrast between long and short for mid and high vowels before /r/. This is the key: The long vowels before /r/ could just as well be categorized as short. Since there is no contrast between /iy, i/, /ey, e/, /uw, u/ before tautosyllabic /r/, we are free to represent fear, fair, boor, etc., without filling the glide position, which otherwise is used to represent the long-short contrast. This in turn

\textsuperscript{30}cf. the split within the word horror ([hɔrə]).

\textsuperscript{31}This merger did not occur in certain Celtic dialects, where /r/ was not a continuant and thus remained outside the Nucleus-Glide positions.
allows /r/ to occur in the glide position in these words.

If /r/ is located in the glide position, the three degrees of vowel length are collapsed into two. The distinction between long and short vowels in most environments remains the same: short vowels have only a nucleus, while long vowels consist of both a nucleus and a glide. But the distinction of long vs. longer vowels preceding /r/ is now changed; it is now the same as the long-short distinction. That is, the nucleus in nucleus-r sequences (now represented as /ir, er, or, ur, ar/, as in beer, bear, bore, boor, bar, respectively) is just one element long, while the vowel in a sequence of nucleus-glide-$-/r/$ is two elements long (where $ represents a syllable boundary). The phonetic tenseness of the nuclei of nucleus-r sequences can be attributed to the vowel-before-vowel raising rule, as discussed below.

In the most thoroughly studied linguistic variable in English (Labov 1966, 1968, Fasold 1969, 1972, Guy 1980, Santa Ana 1991, Patrick in progress, among others) the deletion of /t, d/, preceding /r/ was found to have the same phonetic effect as a preceding vowel. Among various phonetic and grammatical effects, it is generally found that preceding consonants favor deletion, while vowels disfavor it. In most dialects where it was studied, /r/ is generally found to behave like a vowel rather than as a consonant. This is perhaps the strongest of empirical support for this analysis, since it comes from numerous large-scale studies of actual speech in various English dialects. The form of this supporting argument, it should be noted, is rather unusual: the statistical structure of phonetic observations is, I argue, explained in this case by the underlying phonological structure.

To summarize: First, it was shown that though Trager and Bloch place /r/ in post-glide position, there is no need to do so. Second, in just those cases where a glide is unambiguously present (/ay, oy, aw/), /r/ cannot occur within the same syllable; this is explained if /r/ is a glide, and only one glide can occur in a syllable. Conversely, just in those cases where the glide can be considered absent (as in /ir, er, or, ur/, where there is no long-short contrast because of the earlier merger of the vowels in fir, fur, her, and unstressed for, /r/ can follow the vowel nucleus within a single syllable. Thus various independent observations support the theory that /r/ is a glide.

---

32The exception is the Los Angeles Chicano community, where Santa Ana found that /r/ behaves like a consonant in its effects on /t, d/ deletion, an unusual and striking fact that may be attributable to the influence of Spanish contact with that English dialect.

33In Reference American of these contain [3].
3.3.4 The Vowel Within Syllable Structure

Can the concept of “vowel” be defined on syllable structure? (2) above constitutes a theory of the structure of the vocalic part of the syllable: each of the three classes, nucleus, glide, and /r/ fall in a separate slot in the syllable. However, we have now replaced (2) by the simpler scheme in (5):

(5) Nucleus (Glide)

This theory is not immediately compatible with the standard view of syllable structure, in which the rhyme consists of a nucleus preceding a coda. The nucleus is an accepted syllabic constituent, with the same name. But the glide position per se is not an accepted constituent of the syllable. Consider the widely accepted syllable structure in (a).

(a) syll
   / \
  ons rhyme
   / \
 nuc coda

Postnuclear glides, which are not explicitly located in this structure, must occur somewhere within the rhyme. I have shown that glides must occur in a single, non-branching location within the syllable. The fact that it is non-branching may be emphasized by calling this location the glide “slot”. This may be easily expressed as the claim that the glide slot is a labelled, non-branching constituent within the English syllable. The glide node cannot be identical with the coda, because it is non-branching, while the coda can contain multiple segments, in English.\(^{34}\) Nor can it be identical with the nucleus, which precedes it. Therefore the glide node must be within the nucleus as in (b) or within the coda as in (c); or conceivably as a daughter of the rhyme node itself, as in (d).

Various ways of renaming the nodes in (b) and (c) are possible, but the labels are not material to the arguments below. Names like peak, satellite, head, tail, vowel, etc., are available. In (b), for example, I would like to relabel the node nuc’ as vowel, since this would give a maximally simple formal specification of the target of study of this thesis:

\(^{34}\)The argument here must be restricted to English alone, since there are languages in which the glide is the only segment which occupies the coda.
Vowels are nucleus-glide combinations. Others may prefer to relabel \textit{nuc} as \textit{peak}, and \textit{nuc'} as \textit{nucleus}. In (c), the node \textit{coda} could be renamed \textit{tail}, and \textit{coda'} renamed \textit{coda}. The “X-bar”-like notation used here is intended to highlight the arbitrariness of node-labelling.

Which of the structures (b, c, d) is correct? This is an important issue in phonological theory. Since stating the generalizations found in this chapter is somewhat simpler using (b) than with (c), I will examine some relevant arguments below. However, the main arguments of this chapter could be restated in any of the three forms.

Goldsmith (1990:109) states that arguments for a branching nucleus constituent within the syllable are unconvincing. If so, then, we should prefer (c) or (d). The flat rhyme structure in (d) is an unhappy compromise. (d) would allow statements in which glide and nucleus are grouped together to have equal status with those in which glide and coda are grouped. If there were some large number of rules that refer to glide and coda together and an equally large number of rules that refer to nucleus and glide together, economy will still be served if one of the pairs is put into a single constituent. Then at least half of the references in these rules will be simplified, which is a definite improvement over the flat structure, (d), in which none of them would be simplified. In short, the unhappy compromise of (d) amounts to lost generalizations relative to either (b) or (c), and is thus inferior. Next I propose some arguments in favor of the branching-nucleus, (b).

What is the structural significance of the branching, tree-like structure of the syllable?

The application of constituent-structure analysis to syllabic constituents suggests that the tree structures within syllables should have the same structural significance as syntactic tree structures. If the trees in syllable-structure have the properties of context-free grammars, then constraints across constituents, which amount to context-sensitivity, should be impossible. The definition of “context-free” is that elements in distinct constituents cannot
influence each other, while elements within a constituent may. Syllable structure certainly is not perfectly context-free, since there do exist constraints between onsets and nuclei, between nuclei and codas (e.g., the pen/pin merger in the Southern U.S. See also Borowski 1988), etc. Nevertheless, to the extent that syllable-structure is context-free, combinatorial constraints within the constituents of the syllable are acceptable, but across constituents, they are not.

Severe restrictions on combination of vowels and glide are common in the world’s languages (cf. Maddieson 1984). On the other hand, vowels typically combine quite freely with other post-nuclear consonants. Since the context-free grammar view of syllable structure restricts combinations of elements within constituents rather than across constituents, the existence of pervasive restrictions on nucleus-glide combinations, along with the rarity of restrictions on nucleus-coda combinations, would be evidence that nucleus and glides form a constituent distinct from the nucleus-coda constituent (the rhyme). This suggests alternative (b) above.

Another argument for (b) comes from historical phonology. The vowel classes that generally act as a unit in historical change are units formed from the nucleus and glide together. Jamaican Creole provides a number of mergers in which nucleus-glide sequences underwent changes as a unit, as may be seen in transcriptions provided in Wells (1982, vol. 3, p. 576). In Jamaican Creole, a number of Wells’ lexical sets merged into one, while their counterparts without the rhotic glide did not merge. Thus NEAR and SQUARE (/ir, er/) merge, while FLEECE and FACE (/iy, ey/) remain separate, as do KIT and DRESS (/i, e/). Similarly CURE and FORCE (/ur, or/) merged, with no effect on their counterparts GOOSE, GOAT (/uw, ow/) or FOOT, STRUT (/u, a/).

The low vowels (including /oy/) merged into three classes according to their glides: front-gliding, back-gliding, and neither. Thus the nucleus-glide sequence of the low, front-gliding diphthongs of PRICE, CHOICE, (/ay, oy/) merged into JCE /ai/; the long or /i/-gliding vowels of THOUGHT, CLOT, PALM, BATH, NORTH, and START (/ɔ:, ɔ:, a:, ɔr, ar/, all merged into JCE /a:/, and MOUTH (/əw/), the only back-gliding low vowel, remained distinct. Besides the mergers of LOT and TRAP (/a, æ/) and of the unstressed vowels in LETTER, COMMA (/ɔ, ə/), this amounts to a fairly thorough description of the phonological changes in the genesis of Jamaica Creole.
We have seen that it is the nucleus and glide together which define the units which undergo the merger, and which define the units which result from the mergers. In this and many other examples, the nucleus and the glide act together as a unit in historical change. Stating these changes is made simpler when these are treated as a single unit, as in structure (b) above.

Finally, the nucleus and glide positions contain segments which are quite similar in sonority and in phonetic and phonological substance (cf. page 93), while the coda position is less sonorous and distinguishes different kinds of features.

In order to retain the structure in (c), we must reject fundamental assumptions about the significance of branching structure in the syllable: it is nothing like context-free grammar. We must accept increased complexity in historical phonological description. And we must believe the counter-intuitive claim that elements in distinct constituents may be phonetically more similar to each other than elements within a single constituent. For these reasons, despite Goldsmith's claim to the contrary, the structure in (b) seems superior.

3.3.5 Syllable Weight

This section further sharpens our understanding about vowels within syllable structure. The widely observed fact that the "short" vowels cannot occur in stressed, unchecked syllables amounts to the restriction that if the syllable contains a short vowel, a coda must be present. The converse point is also justified by observation. The long mid and high vowels can occur with following /r/ in the same syllable as well as with /r/ in a subsequent syllable, as in Table 3.3 above. Further, these vowels may precede /r/ and other consonants simultaneously: fierce, fort, etc. A coda is not necessary after a long vowel (as in see, say, sigh, Sue, sew, sow, soy). Representing length by the presence of a glide position, we may put these two observations into a single sentence: Stressed syllables must include a coda or a glide. We may even omit the disjunction and reduce it further, as in (6).

(6) Stressed rhymes branch.

35[ʃiː], [ləʊ].
36Phonetically, [siː], [seɪ], [sdʒ], [suː], [soʊ], [səʊ], respectively.

64
If a branching rhyme is understood as a rhyme with a branch anywhere within it, then this statement holds equally under syllable structures (b), (c), or (d).

Syllable weight may be defined as branching structure within the rhyme, so that morae — the units of syllable weight — may be counted by counting the branches within the rhyme and adding one. Then, stress is attracted to heavy syllables. Many languages make a similar distinction between heavy and light syllables, in which stressed syllables must be heavy.\(^{37}\)

This section presented a number of classes of vowel length and complexity, attributing some distinctions to phonetic and suprasyllabic factors, and others to the phonological structure of the syllable. The distributional pattern of “short” vs. “long” vowels is explained by the constraint, Stressed rhymes branch. The low diphthongs were argued to be “longer” even than the “long” vowels because they don’t allow following /r/ in the same syllable. An exemplar of pregenerative English vowel theory in which /r/ follows glides within the syllable was examined and its catalog of counterexamples to the theory that /r/ is a glide was systematically destroyed. We conclude with a theory of the temporal structure of vowels: A vowel consists of a nucleus followed in some cases by a glide. If structure (b) above is assumed, then the concept, “vowel”, may be defined as the constituent containing the nucleus and the glide, there labelled \(\text{nuc}'\). This may be reduced to the context-free-grammar rewrite rule in (7).

\[(7) \text{Vowel} \rightarrow \text{Nucleus (Glide)}\]

This rule, characterizing part of English syllable structure, provides a formal, technical definition of “English vowel”.

### 3.3.6 Implications of the Glide Slot for Dialectal Variation and Historical Change.

The theory of the glide slot has important and clear implications for sound structure and sound change in English history and dialectology. Changes in the inventory or phonetic quality of segments in the glide slot result in a small set of possible readjustments: merger and breaking. Glides or nuclei or both may merge, or syllable-boundaries may be inserted.\(^{37}\)

\(^{37}\)Bailey (1985) suggests that English is one of these.
A number of important phonetic changes are particularly relevant in this context. One is the ancient vocalization of /r/ discussed above, a 17th century English change which is now widespread in English dialects, in which /r/ changes from an obstruent to a continuant. A second is the more recent “r-loss” of Southern British, Boston, some Southern U.S. dialects, and New York City, in which /r/ changed from one kind of continuant (retroflex, or “sulcal” [ʂ]) to another (non-rhotic [o] or [ɔ]). Labov (1966) analyzing NYC desulcalization, called it /r/-vocalization, under the assumption that /r/ is a consonant. A third important phonetic change is the present-day vocalization of /l/, which is developing or developed in various American and British dialects, such as Cockney (Sivertson 1960), Philadelphia (Ash 1982), Pittsburgh and other Western U.S. dialects, etc.

These changes are important, since when a phonetic change occurs so that a post-vocalic consonant becomes vocalic, and is phonologically re-analysed as a glide, all the contrasts among vowels that precede the new glide, which depended formerly on the presence of glide features to distinguish them, must either be re-analysed, or lost, depending on the phonetic forms of the relevant sound classes. Thus there are severe consequences for vowel structure when one of these changes occurs.

First consider the change of /r/ to a vocalic segment. Where post-nuclear /r/ is not a glide, as in those Celtic dialects where the reflex of /r/ is a tap or other non-continuant, there should be no merger among the short vowels: bird, herd, word, and curd should not rhyme. I believe this is correct. As discussed above, where post-nuclear /r/ is a glide, the high and mid short vowels merge before /r/ into a single class of rhotic monophthongs, without distinctive height or backness in the nucleus. In these — most — English dialects, bird, etc. do rhyme. Thus it was the phonetic change of /r/ from a consonant to a vowel that was the force behind the merger of earlier short /ir, ur, er, or/. Note that the short low vowels do not merge (as in car [kaθ], carry [kærθi]). These apparent exceptions are consistent, on the one hand, with the claim that these changes are phonetically rather than phonologically driven, and on the other hand, with the fact that these low vowels, while phonologically short, are phonetically relatively long, due to the larger mouth-opening gesture required for their production.

The result of this phonetic change of /r/ for the “true diphthongs”, /oy, aw, oy/ was breaking. All the segments were retained, but in order to allow the two glides to occur in
a row, Nucleus+Glide+r sequences were spread over two syllables rather than one. This is again consistent with the glide-slot theory, since at most one glide can occur in each syllable.

In dialects where the glide is lost from /ay, aw, oy/, it is predicted that tautosyllabic /r/ should again be possible, so that flour, dire etc., may contain just one syllable. This prediction is successful in monophthongizing Southern U.S. dialects, where these can indeed be monosyllables.

In short, the theory of the glide slot neatly explains one of the most important mergers in the history of English, as well as with phonetic facts in dialects which do not have this merger, as well as syllable-counting facts in dialects which monophthongize /ay/ or /aw/.

The second change, the de-rhoticization of /r/ in dialects like New York City, was explored in detail by Labov (1966). When /r/ becomes an inglide, for example, merger occurs between vowels containing similar nuclei and either inglides or /r/. Thus the vowel in source [soʊs] merges with that in sauce [soʊs], and bared merges with bad (as [beəd]), etc. Again phonetic change in the glide position has major phonological consequences.

The third example of such a change is the ongoing vocalization of syllable-final /l/ in various English dialects. This change goes to an extreme in Pennsylvania, both in Philadelphia (Ash 1982) and Pittsburgh, but also occurs in some British dialects (e.g., Cockney, cf. Sivertson 1960), and to some extent in Western U.S. dialects. The empirical consequences of this change have not been fully described, but the known facts are consistent with the theory of the glide slot. Pittsburgh has extreme /l/-vocalization, where the realization of /l/ is a high-back glide. The result is in many cases a merger of post-vocalic glides with /l/: aisle, owl, Al, and Ow are all pronounced identically as [aʌ]. At the same time, the long-short distinction between certain pairs of vowels may be weakening: the stereotypical example is the local pronunciation of the name of their professional football team, the Pittsburgh Steelers, as the “Stillers”.

Similarly, in the Western U.S., two studies have found mergers of long and short vowels before /l/. Labov, Yaeger and Steiner (1972: 236ff) studied the merger of fool and full in Albuquerque, and DePaolo (1988) documented mergers of certain pairs of long and short vowels before /l/ in Salt Lake City. Unfortunately these studies do not state what the quality of the /l/ is in these dialects. My impression of a couple of speakers (namely,
myself and another speaker) I have analysed informally, from Southern California, a closely related Western U.S. dialect, is that some vocalization of post-vocalic /l/ does occur. Also, I count two syllables in oil and in Owl. These are consistent with a role for /l/-vocalization in the mergers found by LYS and by DePaolo.

The other result of /l/-vocalization which is consistent with the theory is the breaking of sequences of long or gliding vowels plus vocalized /l/ into two syllables. Thus owl could be pronounced [aw$y]; also, "feel", "fail", and possibly even "fool", "foal" could become bisyllabic. The phonetic back-round glide on the long vowels in "fool", "foal", "fowl" may be indistinguishable from the back-unround glide which results from /l/ vocalization, so these may be more likely to undergo glide coalescence rather than breaking into two syllables.38

These facts are tantalizing, and provide support for the theory that a glide slot exists in English syllables. The logic of the argument is as follows. There are four conceptually unrelated changes that seem to occur: (1) the phonetic realization of tautosyllabic /l/ changes into a back vocalic segment; (2) the short-long distinction between certain pairs of vowels is lost through merger; (3) various post-vocalic glides undergo coalescence, or merger with post-vocalic /l/; (4) some syllables with post-vocalic /l/ undergo breaking into two syllables. These changes would seem to be quite unrelated, and the fact that they might be found together might be thought of as a coincidence. However, when considered in terms of the theory of the glide slot, all these changes can be seen as consequences of a single change. The formal description of the change, according to this logic, is that /l/ shifts into the glide slot from its former coda position. The glide slot does not license the [lateral] feature (cf. page 93 below), so it is lost; this accounts for (1), the phonetic vocalization (by which the realization of /l/ is no longer phonetically [lateral]). At the same time, the Glide slot into which the /l/ has moved had formerly represented the long-short distinction; if /l/ now occupies the post-vocalic Glide slot next to formerly long as well as short nuclei, the formal distinction between the long and short vowels disappears; this accounts for the long/short mergers, (2). The same logic accounts for the coalescence of glides with the realization of /l/, (3). Both /-w/ and /l/ cannot at the same time occupy the same slot and also be temporally distinct from one another. Also, as back, [±round],

38Thanks to David Graff and to William Labov for pointing this out to me.
vocalic glides, they are also phonetically quite similar. The result of coalescence of /l/ with /-w/ is explained if /l/ moves into the slot occupied by /-w/, and is phonetically indistinguishable from /-w/. In cases in which the features of the glide which precedes /l/ remain temporally and phonetically distinct from those of /l/, a sequence of two glides results, and the restriction in the theory of the glide slot, that states that only one glide may occur in a syllable, predicts that a sequence of two syllables will result, that is, that observation (4), syllable-breaking, will occur.

Thus these quite different changes may be seen as one. The fact that they seem to co-occur in known cases is support for any theory that explains why they should do so, as the theory of the glide slot does. These very different changes are explained as deriving from a single actual change, which in combination with the theory of the glide slot makes sense of all four otherwise unrelated outcomes of /l/-vocalization that were noted above.

Much research remains to be done to establish whether or not the merger of long and short vowels before /l/ is precisely correlated with the phonetic vocalization of /l/, to chart the course of the mergers as they occur (for example, is there an intermediate stage, where the /l/ is entirely vocalic, but has not been analysed phonologically as a glide?), to explain why some pairs may merge and others may not. An important question is, Why should it be that when /r/ vocalized, the short vowels lost their nuclear qualities, becoming /σ/, but as /l/ vocalizes, the short vowels retain their nuclear qualities and their (back) glides merge into the /I/. A descriptive study of the social, geographic, phonetic, and phonological factors influencing these changes is called for.

In this section I have defined the temporal structure of English vowels; in fact I gave a technical definition for “Vowel”, a certain temporal sequence that is a constituent within the syllable. Let us next consider what are the static components within the sequence, that is, the features which may fill the nucleus and glide positions.

---

39 Part of the answer may be that languages cannot distinguish [±round] back glides.
3.4 Static Vowel Structure

Static vowel structure in English must distinguish three levels of height, and two levels in the front-back dimension. This section discusses various approaches to the formal representation of this three-by-two grid, which has exercised phonologists for many years.

I discuss several types of featural analyses for vowel height. The first, as in SPE, describes vowel qualities in terms of their opposition to a language-particular "neutral position". Taking its language-particularity seriously suggests that the neutral position for each dialect determines its particular feature system. In the standard generativist treatment, vowel height is considered to be multi-dimensional (that is, more than one feature is used to distinguish among heights). This type of treatment is also characteristic of some provocative recent work by Pettersson and Wood (1987a,b; also Wood and Pettersson 1988). In the second type of treatment, vowel height is a multiply-valued single dimension, and various sub-theories (tacitly, theories of counting or numbers) are proposed in order to deal with the multiple values of the dimension. In a third type of analysis, proposed here, privative height features [high] and [low] are used in order to avoid both the combinatorical problems of the standard treatment, and the admission of many-valued dimensions into phonological structure. Finally, the feature [front] is proposed to replace the feature [back].

3.4.1 The Neutral Position

In *The Sound Pattern of English*, Chomsky and Halle's approach to structuring the 3X2 grid began with the observation that, in preparation for speaking, as opposed to breathing or eating, the vocal tract assumes a certain "neutral position." They point out previous claims that this position, the "base of articulation" of Sievers (1901), varies across languages, and claim that for English, it is mid and front, as in [s]. SPE proceeds to define the 3X2 vowel grid on the basis of movements of the tongue body away from this position: up [high], down [low] and back [back]. These dimensions have been considered to be fundamental and universal phonological dimensions for vowel quality, despite the clear implication that the neutral position, and thus the movements of the tongue away from the neutral position, is language-particular.

Not much research has followed upon these observations. However, a major empirical
result of the phonetic investigations in later chapters would seem to be quite relevant, namely, the pattern of phonetic effects of phrasal stress on vowel quality. In most dialects, where a clear pattern is evident, the effect of stress is a graded shift of the phonetic quality of the nucleus of unstressed vowels from the quality of stressed nuclei towards a single "reduction target", which differs from dialect to dialect. This reduction target may be identical to the average location of /o/, the vowel which is phonologically unspecified for quality. These facts are consistent with the identification of SPE's "neutral position", Sievers' "base of articulation", the phonetic realization of /a/ (which is, in Reference American, the unstressed allophone of /ʌ/), and the "reduction target".

Phonological vowel quality features may be derived, as in SPE, by their opposition to the qualities of the reduction target. Thus the phonologically unspecified or unmarked height is that phonetic height which unstressed vowels reduce towards, so that in dialects where unstressed vowels shift towards high-central, the unspecified or default vowel height should be phonetically high (and the unspecified level on the front/back scale should be central). The three American English dialects studied here, Chicago White English, Alabama English, and L.A. Chicano English have a high (central or front) reduction target. (A mid-central reduction target occurs only in Jamaican Creole.) These facts suggest that the neutral position for English dialects is not mid-front.

Either the phonetic foundation for the vowel-quality features, [high], [low], and [back], is to be abandoned, and they are to be stipulated without this justification, or a different set of features may be established - indeed, feature sets, which differ from language to language. For example, if a language had a high-back neutral position as the base of articulation from which sounds deviate, then the marked features [front], and [mid] and [low] or some equivalent formulation, would be indicated.

3.4.2 Vowel Height as Multi-Dimensional.

Summarizing work descending from SPE, Keating (1988:23) provides a "consensus set of segmental features" which, she believes, "in most respects represents current practice". Binary features in this set which are used to describe vowels include those in (8).

(8) [round], [high], [low], [back], [tense], [ATR].
These binary features may have values + and -. The grammar for static vowel structure is the free combination of independent features. This grammar, applied to the height features \([\pm \text{high}], [\pm \text{low}]\) generates four distinct phonological heights, \([+\text{high},-\text{low}]\) (which is interpreted as high), \([-\text{high},-\text{low}]\) (interpreted as mid), and \([-\text{high},+\text{low}]\) (interpreted as low), and a fourth combination, \([+\text{high},+\text{low}]\). This fourth combination is ruled out in SPE by the physical impossibility of the tongue body being both raised and lowered at the same time.\(^{40}\) The form of this argument is typical in modularist argumentation: principles from one module or level of structure are used to rule out structures that are entirely acceptable to other modules. In this way, structure at one level (the combination of \([+\text{high}]\) and \([+\text{low}]\)) is ruled out due to constraints at another level (here, physical constraints on the tongue).

Phonetics and phonology are the most isomorphic parts of grammar. (Surface) phonological structures map directly onto phonetic form, and those features which occur at both levels are argued to have essentially similar content. In no other area of grammar are modules so closely matched. The phonetic impossibility of the phonologically possible combination \([+\text{high},+\text{low}]\) is an example of a mismatch between the two modules: one module accepts a structure, another rejects it. If we prefer linguistic modules to be orthogonal, that is, to be maximally different from each other, then this mismatch is not a problem. If on the other hand isomorphism is preferable between phonological and phonetic structure, it is undesirable. I take the latter view. Phonology is the category-based, abstract mental representation of phonetic form in language, and the isomorphism of the two levels is one of phonology's fundamental properties.

I therefore assume that if phonological features have mutually exclusive values, and thus if a grammar in principle overgenerates, then it is flawed.\(^{41}\) A number of other theorists have been motivated by this view of the problem of vowel height to develop static theories of vowel features which do not have the problems of overgeneration and non-independence of dimensions that the standard view suffers from. Three classes of solutions are given,

\(^{40}\) However, Krohn (1969) and Wang have argued that diphthongs are \([+\text{high},+\text{low}]\).

\(^{41}\) Grammars may incidentally overgenerate, without being flawed. For example Russian requires the \([\pm \text{voice}]\) dimension to distinguish /p/ from /b/, and other features for /X/, and /ɛ/. But /γ/ and /j/ are not distinctive, though /γ/, /j/ do occur (Halle 1957, 1959). The feature system can hardly be at fault for this kind of incidental gap in the system.
those of Wood and his coworkers, and various numbering systems, and a solution I propose that combines privative features and autosegmental tiers. The superiority of any one of these systems over the others in the analysis of English is not clear.

The standard model of vowel quality as defined by the dimensions of height and frontness originated with Alexander Melville Bell’s Visible Speech (1867). Work over the last decade or so by Wood and his coworkers has challenged this classical model on both phonetic and phonological grounds.

In a treatment of the phonology of vowel raising in Bulgarian, Pettersson and Wood (1987) argue that the rule of vowel reduction, in which the vowels /e, a, o/ are raised to /i, ə, u/, respectively, cannot be treated elegantly in a standard binary feature treatment of vowel heights. /e, o/ are mid vowels, and undergo raising to high, while /a/ is a low vowel, and raises to mid. They argue that the standard feature systems are unable to state the raising process in a single rule, since raising must separately change [+low] to [−low], and [−high] to [+high]. They then use articulatory data to argue for the separation of the feature of jaw-opening from the feature of lingual constriction. /e, a, o/ are [−open], while /i, ə, u/ are [−open], so that the reduction rule can be stated simply as: [−open] → [−open] /[−stress]. “This is not the same as saying nonhigh becomes high, since /a/ has the same low pharyngeal tongue posture as /a/. The combination of low pharyngeal tongue posture with close jaw position cannot be expressed by Bell’s features, since a vowel cannot be simultaneously be [+high] and [+low]. It is necessary to distinguish between mandibular and lingual maneuvers. ...” (p. 242)

We may infer that Wood’s analysis resolves vowel height into two articulatory features, such as [open] for the jaw, and [raised] for the tongue, where /i, u/ are [−open, +raised]; /a/ is [−open, −raised]; /e, o/ are [−open, +raised], and /a/ is [−open, −raised]. [Mid] might be defined as a cover feature for [α open, α raised]. Thus all four combinations of two height features do occur, and so Wood has avoided the impossible-combination problem, using a truly multi-dimensional treatment of vowel height. In common practice vowel height is inconsistently held to be both two dimensions (in the formal structure) and one dimension (in phonetic reality). Wood has commendably gotten beyond this inconsistency, since in

---

42 The speech pathologist, better known as the father of Alexander Graham Bell, the inventor of the telephone.
43 For a discussion of the history of this model see Wood (1987).
his view the two dimensions of vowel height are two separate dimensions in articulatory phonetics as well as in the abstract phonology.

3.4.3 Vowel Height as a Single, Numbered Dimension

Many other treatments of vowel height take the opposing view, that height is a single phonological dimension. Such treatments can straightforwardly account for processes which increase or decrease vowel height by one step, as for example NYC English raising of tense low vowels /æh, əh/ (as in man and thought), or Jamaican Creole English raising of long mid vowels /ɛ:, ə:/ (FACE, GOAT) (cf. also processes referred to in Lindau 1978).

If vowel height is a single dimension, then there are many ways to distinguish the points on the scale, all of which amount to different approaches to a theory of numbers. Some (e.g., Ladefoged 1982:262-263; Labov, Yaeger, & Steiner 1972:167ff) use the traditional numbers, 1,2,3,4,..., of base n, where n is greater than the number of points on the scale.

Integers written in any base may be converted to base-2 numbers. Since generative phonologists prefer binary features, they have made a number of attempts to rewrite the familiar decimal scale in terms of various binary numbering systems. In all of these, various combinations of homogeneous binary features — which amount to binary digits — are used to construct a number theory for the vowel-height dimension.

In Clements’ (1989) treatment of vowel height, binary features are used in an interesting way. There, the first binary feature distinguishes among low and non-low vowels; the second among the non-low vowels, etc. Thus the aperture (or height) dimension is organized by dividing the higher height into two. (i e æ) becomes ((i e) æ); (i e e æ) becomes (((i e) e) æ). The second [open] feature cannot occur in a given language unless the first does. The second is added in order to divide the non-low vowels into two groups; the third is added to divide the higher of the non-low classes into two groups. This can be done in a language-particular way, so that if two heights pattern together, the nth [open] feature can be used to distinguish them; while the remaining n-1 [open] features classify them together.

Like other numbering systems, Clements’ [open] features “stack”, so that the nth [open] feature can only occur if the n-1th [open] feature occurs. But unlike Schane’s, discussed
below, they stack inwards, not at the end. This may be clearer if we convert the representations to binary numbers. Clements' system of aperture features may be represented as binary numbers with n binary digits (technically termed bits) representing the n [+open] features. The set of resulting binary numbers are a restricted subset of the possible binary numbers that can be distinguished using n bits.

Considering 0 as [-open] and 1 as [+open], a set of 3 heights are represented as (0., 10, 11); 4 heights are (0., 10., 110., 111), 5 heights are (0..., 10..., 110..., 1110..., 1111). The periods represent bits which are filled in as 0's by a special redundancy rule. After the application of this redundancy rule to a 4-height set of vowels, the resulting binary numbers are those in Table 3.6.

Table 3.6: Numerical representation of height, #1.

<table>
<thead>
<tr>
<th>vowel</th>
<th>binary #’s</th>
<th>Decimal #’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>000</td>
<td>0</td>
</tr>
<tr>
<td>e</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>ε</td>
<td>110</td>
<td>6</td>
</tr>
<tr>
<td>æ</td>
<td>111</td>
<td>7</td>
</tr>
</tbody>
</table>

Further redundancy remains to be factored out, since all digits can be predicted from a single feature specification: the location of the first 0, or, conversely, the location of the first 1. Whichever system is chosen, all other feature-values can be predicted from a single feature. In this system, a given language has a fixed number of these features, or bits; each segment is specified (on the surface) for all values of all the features.

How might a system of privative features work? If a feature were absent, i.e., if the bit were zero, then it would be absent. This is the case in the written representation of normal arithmetic numbers, where leading zeroes are not written.

In fact, a system with these properties has been proposed. In Schane’s particle phonology (1984) the privative, additive, aperture feature, called a, may be thought of as a binary digit in just this way. The front vowels [i, e, ε, æ], are represented as i, ia, iaa, iaaa. Abstracting the height dimension and considering the aperture particles as binary digits, the numbers are as in Table 3.7.
Table 3.7: Numerical representation of height, #2.

<table>
<thead>
<tr>
<th>vowel</th>
<th>particles</th>
<th>binary #’s</th>
<th>Decimal #’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>i</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>e</td>
<td>ia</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>œ</td>
<td>iaa</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>æ</td>
<td>iaaa</td>
<td>111</td>
<td>7</td>
</tr>
</tbody>
</table>

Thus each added aperture particle adds another power of two to the number representing vowel height. Schane's particles are added to the end of the stack of aperture particles, but they could just as well be added at any place in the middle, thus accounting for the language-particular differences which Clements represents by inserting digits at various places.

All of these treatments of vowel height remain controversial. However, as shown here, they can all be seen as attempted theories of numbers. While this seems on the face of it to be rather remote from phonological theory, phonological feature theory can itself be seen as characterizing an inventory of named bits, with particular combinatorial restrictions and phonetic interpretations, that together determine a set of possible numbers, which represent the phonological segments of language.

Next, I will instead present a novel formulation that is different in kind from the above systems, that also solves the impossible-combination problem of SPE's features [+high], [+low].

3.4.4 Privative Height Features

Another solution to the problem of ruling out the formal combination [+high,+low] has two steps. First, consider the features [high] and [low] as privative rather than binary (equipollent) features. That is, they may be present or absent, rather than obligatorily present with one of two values. Second, assume that height is a single dimension. This may be taken as the assumption that there is an autosegmental tier in the feature geometry which may be called the height tier. At most one of the privative features [high], [low] may occupy this tier.

With these assumptions, the three possible height specifications are [high], [low] and
unspecified. Unspecified height may be phonetically interpreted, quite sensibly, as mid height. These exhaust the possibilities; there is no possible combination corresponding to [+high, +low].

A problem with both n-ary and privative features is that there is no provision for making unique references to feature-values, without further formal machinery. Full use of n-ary features might allow feature specifications like

\[
\begin{align*}
[\text{height} &= 3] \\
[\text{height} &\neq 2] \\
[\text{height} &< 3] \\
[\text{height} &> 1].
\end{align*}
\]

Presumably combinations like [height < 1] must be explicitly ruled out. Or if privative features are used, such as [high] and [low] and [front], then, one must be able to distinguish mid height (unspecified) from any height (also unspecified). For example if a process applies to back vowels (which are unmarked for frontness), and not the front vowels (which are marked for frontness), the expression,

\[X \rightarrow Y \left[\_\right]\]

is to be interpreted such that \(\_\) means both unspecified for height, and positively specified as non-front. But this is not formally distinguishable from a process that applies to both front and back mid vowels, where \(\_\) would be interpreted as positively specified as non-high, non-low, and unspecified for backness. Rules need to be able to distinguish these things.

It is evident that phonological rules must be able to state Boolean conditions on the presence or absence of privative features. A formal way of writing this is to allow the use of the negation symbol, \(\neg\), in the statement of phonological rules, whose interpretation is that the feature marked with \(\neg\) is specified to be absent from forms to which the rule applies. Then a segment that is non-back will be specified in rules as \(\neg[\text{back}]\), while a

\[^{44}\text{Stanley (1967) objected that using an unspecified value for a feature amounts to having three values for it. If [height] is taken as a feature, with values [high height], [0 height] and [low height], then his objection would apply. However, height is a tier, not a feature, and it contains features, not values. Only if tiers are considered to be identical to features is Stanley's objection relevant.}\]
segment that could be front or back has no specification stated in the rule. Note that this is rather different from re-introducing ± values in specifying rule environments or feature structures, since α notations are not allowed. Therefore in order to make references to unmarked features unambiguous in the statement of rules, Boolean conditions on the presence or absence of privative features will be allowed in the statement of rules.

In the following discussion, I will assume that privative features represent underlying phonological vowel qualities.

After this treatment of static vowel phonology, I return to the temporal structure of vowels, describing the ways that the static and temporal structures combine so as to represent the vowel categories of Reference American.

3.5 Reference American Vowel Structure

In this section I present a structural analysis of the surface phonology of English vowels. A system with a five-vowel base (i, e, a, o, u) combined with various glides is discussed and dismissed; then, a “base-6” system (i, e, æ, a, o, u) is examined and various apparent problems and actual virtues are brought forth, including a discussion of the structure of the phonological representation of postvocalic /r/ (cf. Bhat 1974).

It was shown above that RA English syllables contain a glide slot, which is occupied by at most one of /y, w, r/. Economy will be served if some distinctions among English vowels can be attributed to this independently motivated structure instead of to other vocalic features. Therefore the contrasts between Wells’ classes FLEECE and KIT, GOOSE and FOOT, FACE and DRESS, GOAT and STRUT, which could be attributed to a static feature such as [tense], [peripheral], [long], etc., can instead be attributed to the (temporal) structure of the syllable, that is, the presence or absence of a glide slot. Consider the preliminary hypothesis, following Trager & Bloch, that the long vowels contain glides which are identical to the front and back glides of /ay, oy, aw/. Then one may put these vowels, for which the details of the feature-specifications for the glides are unclear, into a structure with the short vowels, the low diphthongs, and the r-glides. I make the basis a 5-vowel system, because the Vy and Vw vowels, joined together, fit into a space with 5 slots. This analysis is represented in Table 3.8.
Table 3.8: Preliminary Reference American vowel structure (base-5).

<table>
<thead>
<tr>
<th>V</th>
<th>Vw</th>
<th>Vr</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>iy</td>
<td>□</td>
</tr>
<tr>
<td>u</td>
<td>uw</td>
<td>ir</td>
</tr>
<tr>
<td>iy</td>
<td>□</td>
<td>ur</td>
</tr>
<tr>
<td>e</td>
<td>ey</td>
<td>□</td>
</tr>
<tr>
<td>a</td>
<td>ay</td>
<td></td>
</tr>
<tr>
<td>ey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>ow</td>
<td>er</td>
</tr>
<tr>
<td>a</td>
<td>aw</td>
<td>or</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Three low vowels cannot be fitted into this system: /æ, ə, a:/ as in TRAP, THOUGHT, and PALM, traditionally named “short A”, “long open O”, and “broad A”, respectively. The unstable status of these low vowels is diachronically evident in that in different dialects they variously undergo raising, breaking, merger with other vowels, etc.

In some dialects (in particular, Reference American), these vowel classes are kept distinct, so it is important to provide a structural representation for them. Considering the set /ə, æ, ə:, a:/ as a group, we may divide them into short and long, front and back, as in Table 3.9.

Table 3.9: Low monophthongs in Reference American.

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>short</td>
<td>æ</td>
<td>a</td>
</tr>
<tr>
<td>long</td>
<td>a:</td>
<td>ə:</td>
</tr>
</tbody>
</table>

The reason /ə:/ is considered to be long, while /a/ is considered short, is that /ə:/ occurs in free position, while /a/ does not. Further, they have historically been analysed as long and short (cf. the traditional names “long open O” and “short O”). Similarly, /æ/ (“short A”) is a checked, therefore short, vowel, and /a:/ is a free, therefore long, vowel.

Given that this length dimension is necessary for the low vowels, the same dimension might also be used for the mid and high vowels as well. In Trager & Bloch’s system represented above in Table 3.5, /h/ signifies both length and ingliding, so the Vh vowels are their (pure) long vowels, which I will signify as V: rather than Vh. The three words idea, yeah, huh and words of those classes must be eliminated if the sets FLEECE, FACE, GOAT are to take their places in the V: sub-system. Idea seems properly classified with words like

---

45Two other lexical sets which go with these are CLOTH, which in the U.S. goes with THOUGHT and BATH, which in the U.S. goes with TRAP.
Maria, Judea, diarrhea, etc., in which the final vowel sequence contains two syllables, and may be written /iy$ə/.

Yeah, huh are among a quite small class of utterances in English which are difficult to analyse phonologically, including the children’s teasing call, [nænænæ'nænæ], the agreement sounds [ʌhʌ, mɪhʌ, ɳɪhɭ, ɳɪhɭ], and the disagreement sounds [ʌfʌ, mɪfɪʌ, ɳɪfɭ, ɳɪfɭ].

Huh is a monosyllabic form similar to the bisyllabic forms in the agreement or disagreement classes. Indeed in at least one meaning it can be replaced by hmm (also acceptable as [hɭ], [hɭ]). These are evidently quite different from fully specified phonological forms, since they lack specification for place of articulation or even oral closure for the syllables. If they freely vary from forms including the default vowel [ʌ] to forms including syllabic nasals at various places of articulation, it would seem that they need not be given a phonological representation in the language.

The phonological form of yeah, if there is one, is difficult to establish. In Philadelphia, I notice that African American speakers say [yi:ð]; I believe I pronounce the word as [yə:]. In the Chicago speech analysed below, while Rita says [yɛ'z] for yes, she has a long glide down the front of vowel space for yeah, which could be suggests either [yə] or [yɛ], but not [yɛ'], which would be expected if it had the vowel in DRESS ([drɛ'z]). If the form is [yə], then it is like the teasing call, [nænænæ'nænæ], in that it contains an utterance-final, stressed [ə], which would seem to be ruled out by the fundamental principle of English phonology, that short vowels (like “short A”, or /æ/) cannot occur in a stressed syllable without a coda. Instead of abandoning basic principles of English phonology to accommodate these marginal forms, we may instead classify yeah with affective forms like those discussed above, which do not have a fully specified, normal, phonological representation in the language.

With idea, yeah, huh now eliminated from Trager & Bloch’s long vowel sub-system, gaps which correspond neatly to the /iy, ey, ow, uw/ classes are left empty. The lexical sets, PALM and THOUGHT, occupy the two low slots in the V: sub-system.

Next, notice that the vowels which may now be written as /i:, e:, ə:, o:, u:/ (as well as lengthened /æ/: in dialects where /æ/ has become long) undergo raising of the nucleus relative to the corresponding short or simple vowels. This does not apply to the nuclei of the “true” diphthongs /aɪ, aw, oy/, which in T&B’s system are classified with the above vowels as containing /-y, -w/. This raising relationship is sometimes referred to
as the vowel-before-vowel raising rule, discussed below, which applies to the nuclei of all these long vowels, but not the true diphthongs, in various dialects. Thus it would be an improvement on T&B’s system to analyse the mid and high vowels as /iː, eː, ɔː, uː/, that is as long vowels similar to /aː, ɔː/, so that this rule may be stated simply, as the raising of long vowels. This would imply that the system has a six vowel base, an implication that is an important prerequisite for any system in which /æ/ is to be analysed as a short vowel along with /i, e, æ, a, u/. I therefore propose as the surface phonological vowel system for Reference American, the “base-6” structure given in Table 3.10.46

Table 3.10: Reference American vowel structure

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>V:</th>
<th>Vr</th>
<th>Vy</th>
<th>Vw</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i</td>
<td>u</td>
<td>iː</td>
<td>uː</td>
<td>□</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>ɔ</td>
<td>eː</td>
<td>ɔː</td>
<td>□</td>
</tr>
<tr>
<td>low</td>
<td>æ</td>
<td>a</td>
<td>æː</td>
<td>aː</td>
<td>□</td>
</tr>
</tbody>
</table>

3.5.1 Apparent Problems in the Base-6 System

There are seven apparent problems with this base-6 system, which I will discuss and overcome in turn.

One apparent problem is the large number of gaps in the Vy and Vw classes. These slots may be labelled /iy, ey, æy, uy, iw, æw, ow, uw/. Several of these gaps seem motivated on phonetic grounds: the slots /iy, ey, ow, uw/ cannot be filled in a manner that would be phonetically distinguishable from the corresponding long vowels /iː, eː, ɔː, uː/, with which Trager & Bloch identified them. This could be taken as an argument that these four vowels are not plain long vowels, but front- and back-gliding vowels, as indeed

---

46 We may locate Wells’ lexical sets in the slots of this structure, thus:

<table>
<thead>
<tr>
<th>V</th>
<th>V:</th>
<th>Vr</th>
<th>Vy</th>
<th>Vw</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT</td>
<td>FOOT</td>
<td>FLEECE</td>
<td>GOOSE</td>
<td>NEAR</td>
</tr>
<tr>
<td>DRESS</td>
<td>STRUT</td>
<td>FACE</td>
<td>GOAT</td>
<td>SQUARE</td>
</tr>
<tr>
<td>TRAP</td>
<td>LOT</td>
<td>PALM</td>
<td>THOUGHT</td>
<td>(marry)</td>
</tr>
</tbody>
</table>

Not displayed are BATH, CLOTH, FORCE, NURSE, and the unstressed sets HAPPY, COMMA, LETTER. BATH goes together with TRAP; CLOTH with THOUGHT; FORCE with NORTH; COMMA with STRUT; LETTER with NURSE; HAPPY with FLEECE (though in the South, HAPPY typically goes with KIT). NURSE is analyzed separately from this system, as argued below.
phonetically they are, in most dialects. However, this would lead to gaps in the V: system for /i:, e, o, u:/, which would then be similarly unexplained. Whichever set of forms are chosen as underlying, the other set of forms would not be phonetically distinguishable; therefore there can be only one set. The choice made here is preferred because the glides are phonetically predictable, because the underlying specification of (plain) long vowels is simpler than for gliding vowels, and because the raising of nuclei of long vowels is more simply stated when /i:, e, o, u:/ are considered to form a natural class of long vowels opposed to /a/, aw, oy/.

The remaining 5 gaps are arguably legitimate, true gaps in the system, assuming that phonological gaps are often a precondition for phonetic vowel-shifting to occur. Thus the slots which may be labelled /uy, æy, iw, æw, ew/ are phonetically filled by vowel shifts in various dialects. Thus /ay, oy/ undergo a chain shift to [o', u'] in various dialects (e.g., Southampton, England). [æ, e] precede a back-glide in the fronting and raising shift of /aw/ in Philadelphia and the Southern U.S. (which in mild form results in [æʌ] and in extreme forms can go as far as [eɔ], merging crown with crayon). [iʌ], which may be phonologically identified as /iw/, can result from the fronting of /uw/ in the speech of many young, white, suburban U.S. speakers. /iw/ has been argued to be the phonological form of the word dew in those dialects which retain the marginal difference between dew and do, which has almost disappeared in the U.S. Finally, [æi] occurs phonetically in the Southern U.S. where /æ/ precedes /ʃ, η/; this shift might have resulted in merger if there weren’t a gap in the position labelled /æy/. Thus these gaps may plausibly be seen as true gaps in the system, which are preconditions for some of the phonetic changes that do indeed occur.

Next, notice the asymmetry between the Vy and Vw paradigms, which are full of gaps, and the Vr paradigm, which has no gaps at all. This asymmetry may be explained by the different historical sources of these sub-systems: The glide /r/ descended historically from a flapped or trilled consonant. Vowels and single coda consonants normally combine quite freely; this was evidently the case for vowel-r sequences in an earlier stage of English. Thus the current vowel system simply inherited this gap-less vowel-r paradigm when /r/ became a glide.

A second problem which both the base-5 and base-6 systems have difficulty with is the
4-way (height) contrast in the set, weary, Mary, merry, marry. Many dialects (such as my own) merge the latter three classes, while other dialects merge various other subsets of the three or change elements in other ways. However, in those dialects where the difference is maintained, at least one of the four must be analysed, in the present approach, as different in temporal structure from the others. The most reasonable possibility is to put the /r/ in Mary into the next syllable, thus allowing Mary to have a tense nucleus. If /r/ is in the second syllable in Mary, the glide position in the first syllable is available to make the vowel long, and thereby phonetically tense. That is, Mary is analysed like eyrie was, in the above discussion of Trager & Bloch. The other vowels in this set would have no such boundary between the nucleus and the /r/. Phonetic analysis of these vowel classes in a dialect with the necessary distinctions (e.g., New Jersey) might support or disconfirm this hypothesis.

The fact that this set of contrasts is difficult to represent in this system might be taken as evidence against this analysis. However, one criterion for judging the costliness of a phonological structure is whether or not it is eliminated through historical changes. The complexity of structural differences should be related to their instability in linguistic history. Marginal distinctions should have only a marginal place made for them in the phonological structure. It would be quite sensible if distinctions which are phonologically costly to represent were also difficult to maintain and historically unstable. If a set of contrasts were simplified in different ways in different, independent historical changes, this would constitute evidence in support of structures in which the contrasts are difficult to represent. This is the case here: the various mergers and sound changes have had the effect of making it possible to represent the remaining contrasts in a simpler way within this system. In dialects like my own, Mary=merry=marry; in other dialects Mary=merry. In Philadelphia merry is nearly merged with Murray. Far from being negative evidence, the facts of these various mergers and sound shifts, which have the effect of simplifying this difficult-to-represent set of distinctions, provide support for this analysis.

Third, the nuclei of /ɔ:/ and /or, oy/ are representationally distinct, despite the fact that for some speakers of dialects with the contrasts, /a/ ≠ /ɔ:/ ≠ /o:/ (with minimal pairs, cot, caught, coat), the nucleus of /or/ (court) or of /oy/ (boy) may be subjectively identified.

---

47See the discussion of near-merger in appendix 2, page 306.
with /ɔː/ rather than /oː/. If speakers' intuitions match their phonological structures, then this structure would not predict this identification. These speakers may have a different structural system, such as the one proposed in a later chapter for Alabama English.

Fourth, mergers and splits may modify the contrasts in this structure. For example, I myself make no distinction among /aː; ɔː; ɑː/: the long low vowels have merged with the short low back vowel. If the above structure is taken as describing my dialect, then it now has gaps, which are not good. But phonemic mergers generally do create gaps in phonological structures. Since mergers nevertheless continue to happen, the unfortunate gaps in this structure — in dialects which merge various phonemes — may simply be unavoidable.

Fifth, /ær/ has a restricted distribution: it doesn't occur word-finally. /ærV/ exists (marry, sparrow, etc.) but /ær#/ does not.⁴⁸ On the other hand, the slot for /ær/ that the base-6 system provides is an improvement on the coverage of the VrV vowels provided by the base-5 system. /æ/ has no slot in the base-5 vowel system, and /ær/ doesn't either. So this is at least an improvement, if not a perfect solution.

Sixth, there is no slot for /ɔr/, distinct in some dialects from both /or/ and /ər/. Wells gives this a separate lexical set, NORTII, as opposed to FORCE and START. In Alabama, all three appear to be kept separate. The structure of the Alabama Vr sub-system is treated at length in that chapter, and found to be compatible with a base-6 analysis. The /ɔr/ and /or/ classes merge in most dialects, either with each other or with other Vr classes. In most American dialects, NORTII is not distinguished from FORCE, while in Jamaica, for example, NORTII is pronounced like START, and FORCE is pronounced like CURE. In the current proposal, this distinction is a difficult one to represent. The various mergers appear as natural, if not predictable, responses of the system to the representational costliness of the distinction between these vowels. Again we find that a distinction that does not fit the paradigm is lost in different dialects and in different ways. Such mergers then appear as simplifications of hard-to-represent distinctions, providing support for the structure rather than evidence against it.

The seventh apparent problem is the representation of the rhotic monophthong, /ər/ ⁴⁸Southern dialects may have [æɔ] in SQUARE, but this may be analysed as lowered /ər/ rather than as /ær/, as discussed in the Alabama chapter. If /ær/ were correct for these dialects, then so much the better for this analysis, since /ær#/ would then fill this gap in the system.
(as in NURSE). If /ə/ is taken as a sequence, /ər/, then presumably the featural content of the /ə/ nucleus is nil. However Reference American already has a vowel with this structure, namely /or/. /or/ is not [high], [low], or [front],\(^{49}\) so its nucleus in this system is unspecified for vowel-quality features; its glide is specified as /r/. Thus /ər/, as /ər/, has the structural representation in Figure 3.2, identical to that of /or/.

Figure 3.2: The underlying representation of /or/.

\[
\begin{array}{cc}
N & G \\
\mid \\
\text{[rhotic]} \\
\end{array}
\]

Some other structural difference must distinguish them. Since [high], [low], [front] are unavailable, the only underlying structures that may be used are the “features” Nucleus and Glide. Both /or/ and /ər/ contain a glide slot, so the remaining possibility is to contrast them by presence vs. absence of the nucleus. Thus, /or/ is in this analysis a vowel consisting (underlyingly) of a nucleus without any features, plus a [rhotic] glide. On the other hand /ər/ is analysed as a vowel consisting of a [rhotic] glide, with no nucleus at all. Later in the derivation of phonetic forms (after the nucleus of /or/ is specified for other features such as [round]), the nucleus slot for /ər/ is inserted by rules which ensure the satisfaction of well-formedness conditions on syllables (such as the condition that requires a nucleus in each syllable; these are part of the process of syllabification). Then the surface vowel-quality features of the nucleus are then inherited from the glide’s features, making /ər/ rhotic throughout. The following derivation (for r-ful dialects) relates the underlying form of syllabic /r/ (written /ər/, and represented as G[rhotic]) to its surface form, a long monophthong that is rhotic throughout.

Indo-European is considered to have underlyingly unspecified vowels, the so-called zero-grade vowels. Yip (1987) has also argued that “suffix-initial /i/ in English is absent underlyingly and is inserted by rule.” These previous analyses lend further plausibility to

\(^{49}\)For the sake of consistency in features, I represent the front-back dimension with a privative feature, either [front] or [back]. The choice between these two is discussed below where the use of the non-traditional feature, [front], instead of the usual feature, [back], is justified.
G N + G N G
[rhotic] 1 [rhotic] 2 [rhotic]

1. Nucleus insertion, to satisfy syllable well-formedness conditions.

2. Spread of vocalic features onto the adjacent unspecified slot.

what might seem a rather unintuitive proposal that /ə/ has no underlying nucleus.

This analysis of /ə/ is quite intriguing on a number of grounds. On the one hand, there is little direct (phonetic) evidence that /ə/ is underlyingly a sequence; /ə/ is a phonetic monophthong in most dialects, excepting Celtic dialects, where /r/ is not a glide at all, and excepting dialects similar in this respect to New York City, where a subclass of /ə/, /əV/ (as in hurry), contains [r]. On the other hand, /ə/ has often been analysed as a sequence of a+r, perhaps because both of these segments are independently motivated in nucleus and glide positions, and because /r/ is not otherwise motivated as a nuclear segment. The present proposal accommodates both views. Since the nucleus links to the vowel-quality features of the glide, both positions have identical phonetic content, thereby formalizing the fact that /ə/ is a monophthong. But we retain the claim that rhotic features are derived from, or underlyingly restricted to, the glide position, since in the underspecified underlying form, the nucleus itself is absent. Thus what appears to be a problem in the representation of /ə/ leads to a tidy solution in which formerly opposing views of the nature of this vowel are simultaneously accomodated.

This analysis of /ə/ also helps us to understand the merger of fir, fur, (trans)fer, and (unstressed) for. These short vowels before /r/ lost their nuclear vowel-quality features, and — in this analysis — lost even the Nucleus specification. This analysis is useful because it explains why it should be that the resulting merged class of /ə/ vowels no longer fits into the base-6 set of nuclei. The nucleus of /ə/ is neither [high] nor [low] nor [front], nor is it even "none of the above" (as is the mid-central vowel, /ʌ/). /ə/ is something else entirely, namely [rhotic]. This analysis makes it explicit and formal that /ə/ is not part of the base-6 system. Since it has no nucleus, it cannot be a part of any system of nuclei.
The seven apparent problems have evaporated, led to plausible and interesting analyses, or metamorphosed into strong support for the analysis. None remain serious enough to compel the abandonment of the base-6 analysis of Reference American proposed here.

3.5.2 Features in Reference American

What is the featural composition of the present system? Statically, there are three vowel heights, low, mid, and high, plus two degrees of backness, front and back, within the nucleus. Additionally, there must be structures which accommodate both length and for gliding.

Treatments of vowel length in the early structuralist literature argued over the relative merits of representing long i, for example, as /i:/ vs. /ii/ vs. /iy/. In more recent phonology (SPE, for example), a static feature was used to distinguish the "short" and "long" vowels, so that what here are long vowels was distinguished from their short counterparts as [peripheral], [tense], [long], or some other feature. In autosegmental theory, there are a number of kinds of structures which are used to represent vowel length and complexity. Where length alone is the concern, the complex of features which represents the vowel's quality becomes linked to multiple timing slots. E.g., a long low vowel, like the "broad A"(/a:/, PALM) of father is contrasted with a short low-front vowel, like "short A" (/æ/, TRAP) of blather by associating the same vowel quality features to two timing slots instead of one. Where vowels are phonetically complex, as in diphthongs, the syllable structure again licenses multiple vocalic timing slots, but instead of linking them to a single static feature-complex, they are separately linked to distinct feature-complexes that specify vowel quality. Thus the features for the /a/ in /ay/ are linked to the timing slot of the syllabic constituent in the syllable (the nucleus), and the /y/ features are linked to the timing slot of the non-syllabic constituent (the glide).

In this proposal, we have argued for the existence of a glide position within the syllable which may be specified for features that distinguish front, back, and rhotic glides. This glide position is absent in syllables containing the short vowels, and the constraint that stressed rhymes branch is met in that coda consonants occur with stressed short vowels. Given this arrangement, which accounts for the V, Vr, Vy, Vw classes, and given the class
of long vowels to be represented in this system, it is unnecessary and undesirable to use some static feature, such as [long], to distinguish them. Instead, it is preferable to use the independently motivated temporal structure, namely the nucleus and glide positions of syllables. This dovetails nicely with autosegmental theory, which deals with vowel length via multiple timing slots.

The proposed structure of Reference American also lends itself to economical representation in terms of features. Three degrees of height and two of backness must be distinguished for nuclei; glides may freely be unspecified (for long vowels) or specified as rhotic, or specified as front or back with combinatory restrictions on the nuclei they may co-occur with (/ay, oy/, /aw/ are the only vowels with underlying front and back glides). A number of featural analyses for this structure are possible, as already discussed above in the section on static vowel features.

The solution proposed here for the problem of [+high,+low] incompatibility is to have [high] and [low] be privative features which occur on a height tier characterizing syllable nuclei. If the feature that distinguishes front from back may also be taken as a privative feature, then we may have a uniform formal treatment of vowel-quality features, which is desirable if it can be maintained.

I make the reasonable proposal that central vowels are phonetically unmarked in the front-back dimension — reasonable, since they lie in the middle of the front-back dimension. In English, the vowels with central nuclei /a, a/ are to be classified with the back vowels, in opposition to the front vowels. That is, there is a full set of front vowels, but the back/rounded and the central vowels are in complementary distribution in the phonological structure, as I will now show.

Consider the pattern of the specification of rounding on the non-front vowels for Reference American. Backness and rounding are related in a hierarchy of both height and vowel complexity: the [round] feature is (redundantly) applied to the high, mid, and low long back vowels (/u:, o:, ɔ:/), to the high and mid back vowels before glides (/ur, or, oy/ but not /ar, ay, aw/), and to the high back short vowel (/u/ but neither of /ʌ, a/). In short, the more complex the sub-system, the more heights have rounding, where long vowels are more complex than vowels with glides. (British English is somewhat different, since the

50See for example, L.A. Chicano, page 259.
short, low vowel here labeled /a/, as in LOT, is [round], as short [ɔ].) This implicational pattern is shown in Table 3.11.

Table 3.11: Implicational constraints on the feature [round]

[round] applies to vowels above the line.

<table>
<thead>
<tr>
<th>Long</th>
<th>Gliding</th>
<th>Short</th>
</tr>
</thead>
<tbody>
<tr>
<td>u:</td>
<td>ur</td>
<td>u</td>
</tr>
<tr>
<td>o:</td>
<td>or oy</td>
<td>a</td>
</tr>
<tr>
<td>a:</td>
<td>ar ay</td>
<td>a</td>
</tr>
</tbody>
</table>

Within the different classes of sub-systems in Table 3.11, back, rounded nuclei occur in complementary distribution with central and unrounded nuclei. Because of this complementary distribution, the back/rounded and the central nuclei of Reference American are underlingly represented with the same value on the front/back dimension, and an implicational rule describing the pattern in Table 3.11 is assumed to be responsible for the redundant marking of the back/round vowels with the features [back] or [round] or [back/round]. Chapter 2 showed that tongue-backing and lip-rounding have the same acoustic effect in the acoustic-phonetic front-back dimension. If acoustic dimensions are taken as the fundamental to vowel quality in English, then backing and rounding are mutually reinforcing articulatory aspects of the same acoustic-phonetic feature.

Chapter 2 also shows that phonetic [ɔ], the resonant sound of a uniform acoustic tube closed at one end, is a reasonable approximation to the average vowel quality for a number of English speakers, and suggests that other vowels are to be understood acoustically as deviations from this average position. Thus the mid-central articulatory position whose sound is written as [ɔ] (this is different from the usually raised position of the realization of unstressed /a/, which may sometimes be written as [ɨ]) is “unmarked” in a phonetic sense as well: it is the center from which other vowels are understood as deviating.

Since the central vowels can therefore be understood as phonetically unmarked, and since the back vowels are in complementary distribution with the central vowels, I will consider the central and back/round vowels to be underlingly unmarked.

The underlingly marked set of vowels on the front-back dimension are thus the front vowels. The privative feature used to make the underlying distinction between the front
and central or back vowels is then assumed to be [front], rather than the traditional [back] feature of SPE, and other generative work. At a lower level, the feature [back] or [round] or [back/round] is added by the assumed implicational rule to a subset of the unmarked vowels.

I thus represent the static structure of Reference American with the three privative features [high], [low], and [front]. This structure is compatible with theories of markedness. The unmarked short vowel is central (i.e., non-front) and mid, namely /a/. The choice of [front] as the marked value for the front-back dimension allows /a/, defined as the unmarked stressed vowel, to be phonologically identified with /a/, as my intuitions and as monitored pronunciations in my dialect suggest. In SPE, on the other hand, where [back] is the front-back feature, the same formal step would lead /a/ to be identified with as the unstressed allophone of /e/, an unlikely result. In this proposal both /a/, and /a/ are phonologically unmarked, which seems quite sensible, and /a/ is the unstressed allophone of /a/ in Reference American. (Note that /a/ = /a/ is not true of British Received Pronunciation. Also, in Southern U.S. dialects which distinguish /a/ and /i/, there are further complications.) In this proposal the cross-linguistically infrequent sound, /æ/, has a marked frontness value, as it should. After the redundant feature [round] is added to the specifications for the relatively higher non-front vowels, then /u; ur, or, oy; u; o; e:/ are all marked for roundness (as discussed above), while /u, a, o/ are unmarked in the front vs. back/round dimension, which, again, makes phonetic sense since they are central vowels.

To summarize, the six vowel types which form the basis of each Nucleus-Glide sub-system have the underlying feature specifications given in Table 3.12.

Table 3.12: Feature specifications for nuclei.

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central/Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>[high,front]</td>
<td>[high]</td>
</tr>
<tr>
<td>Mid</td>
<td>[front]</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>[low,front]</td>
<td>[low]</td>
</tr>
</tbody>
</table>

Thus, three privative features with seven total occurrences in the structure are used to specify six vowel qualities, which is the maximum possible economy here.

Now let us examine glides. The glide position may freely (i.e., in combination with any
of the nuclei) be unspecified (for long vowels) or specified for a [rhotic] feature of some kind (for Vr vowels). The featural representation of the restricted front- and back-glides in /aw/(MOUTH) and /ay, oy/ (PRICE, CHOICE) is the last step. In order to distinguish /ay/ from /aw/ in dialects where they have the same nucleus, we must distinguish front glides from back glides. If we use the same features as are used for the front-back differences among nuclei, then the glides in /ay, oy/ are marked by the feature [front], while the back-glide /aw/ does not need to be specified for backness.

However, if the glide /aw/ is completely unspecified, then there will be no difference between the underlying representations of /ɔ:/ and /aw/. These both have non-front, low nuclei, and both would contain a glide which is completely unspecified. This is, of course, the underlying representation for length: a glide without static features. If the long vowel /ɔ:/ has no features specified in its glide slot, then the diphthong /aw/ must be specified in some way to mark the contrast; we may use the feature [high] for this purpose. The feature-combinations [high,front] and [high] instead of [front] and [ ], are therefore used to distinguish the front and back glides from each other and from the long vowels, which have unspecified glides. Thus the possible underlying glide structures are as in Figure 3.3.

Figure 3.3: Possible glide structures

<table>
<thead>
<tr>
<th>(V:)</th>
<th>(Vr)</th>
<th>(Vw)</th>
<th>(Vy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>[rhotic]</td>
<td>[high]</td>
<td>[high,front]</td>
</tr>
</tbody>
</table>

Because the /-w/ glide is high, it falls in the class of redundantly marked [round] vowels of Table 3.11

Some of these glide structures can occur with all six nucleus values in Reference American, each of them forms what I call, following Labov, a vowel "sub-system", sets of vowels which have important roles in the patterns of misunderstanding (Labov 1990) and sound change (LYS).

Summarizing, the following formal system will generate the possible vowels of Reference American.
(1) Vowel $\rightarrow$ Nucleus (Glide).

(2) Nucleus licenses [front] and the height features [high] and [low].

(3) Glide licenses [front], [high], and [rhotic].

Rhoticity and lowness

Notice that the feature set, [front], [high], [low], is rather similar to the feature set, [front], [high], [rhotic], and also that the features [low] and [rhotic] occur in complementary distribution within the vowel: one occurs only in nucleus position and the other occurs only in glide position. This suggests the proposal that the two features are underlingly just one — [low] — and that the [rhotic] feature which characterizes the phonetic form of these glides may be added later by redundancy rules. This proposal is strengthened by a number of observations: 1) The vowel $[\exists]$ has a non-high tongue-body position; 2) In “r-less” dialects, the reflex of $[\exists]$ is a non-rhotic, non-high inglide; 3) in these dialects (e.g., British Received Pronunciation) the reflexes of Vr sequences with a non-high nucleus are phonetically monophthongs.

Under this proposal, the difference between rhotic and non-rhotic dialects is in whether or not the redundant [rhotic] feature is added to the underlying forms. The underlying forms are identical. Note further that the Vr sequences in r-less dialects that are monophthongs — those with mid (that is, relatively low) nuclei — can be explained naturally if $/r/$ is simply the feature, [low]: the glide and the nucleus are both low, and in a process much like the redundant specification of vowel quality features on the glide in long vowels, the vowel quality features of the nucleus are inherited by the glide, with which it is already entirely compatible. Thus the nucleus and glide merge. (This argues for the specification of [rhotic] preceding the nucleus-to-glide inheritance of other quality features.)

Finally, consider the issue of phonetic implementation of glide features. Glides do not uniformly attain a particular degree of phonetic height. The glide in /ay/ may reach $[\exists]$, while that in /oy/ reaches $[\z]$. /aw/ is commonly $[\alpha^*]$ rather than $[\alpha^*]$. Hence a glide which is [high] is relatively high. This analogy may be applied to r-glides: if they are underlingly [low], then this is relative: high nuclei may glide inward and downward to a phonetic height of mid, rather than low. This makes 1) and 2) compatible with a feature [low] rather than
[-high] or some such other, merely non-high, feature specification: [low] in glide position simply means relatively low.\(^{51}\) (3) above thus becomes:

(4) GLIDE licenses [high], [front], and [low].

(4) in turn may be collapsed with (2) by the two following statements:

(5) Nucleus and Glide are vocalic.

(6) [front], [high], and [low] are vocalic features.

The first of these may be eliminated since it simply restates the conclusion of Section 3.3.5: Vowel → Nucleus (Glide). The second is perhaps the fundamental stipulation of vowel phonology; it might be better stated in terms of phonological licensing: The vowel constituent within the syllable licenses the feature [front] as well as the height tier with privative features [high] and [low].

A later rule redundantly marks [low] postnuclear glides as [rhotic], in r-ful dialects.\(^{52}\) These rules give formal featural representations to the structure in Table 3.10 (page 81), as repeated in Table 3.13. As a convenient linear representation for the multi-linear structure of vowel features occupying Nucleus and Glide slots, I will sometimes use the notations N[...], and G[...] to denote features [...] in the Nucleus and Glide positions in the syllable, respectively. V, V:, Vr, Vv, Vw are by now familiar as names of sub-systems.

This proposal has a number of formal virtues. It is extremely simple in terms of the number and type of underlying features: [high], [low], [front]. It is also has relatively few gaps; none except in Vv and Vw combinations, plus the restriction of /œr/ (that is, N[front, low]+G[low]) to the environment where an unstressed syllable follows. The degree of underspecification is fairly extreme: the total number of features (counting the presence

---

\(^{51}\) Another perspective on this issue is the following. Notice that the glide position is used to encode four things: the three glides /y, w, r/, and length. This requires a front/back distinction, and a high/non-high distinction, but it appears that there is no need for a three-height distinction among glides. The multiple distinctions of height among long vowels may presumably be dealt with at a lower, phonetic level, where the height of the glide is phonetically assimilated to that of the nucleus. If the glide position need not distinguish three underlying degrees of height, then the [low] feature in glide position may be re-interpreted simply as non-high. This would make better sense of the three instances where “non-high” was used in the preceding two paragraphs.

\(^{52}\) r-less dialects retain /r/ in syllable onsets, which may be specified as [rhotic] underlyingly. Where r-less dialects have postvocalic “intrusive” /r/, as in “A vodka or two” [əvəʊkdərəu̯], the intrusive /r/ must be analysed as non-underlying, since it cannot be syllable-initial.
Table 3.13: Feature structure of the Reference American vowel system

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[high]</td>
<td>i</td>
<td>i:</td>
<td>ir</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>[ ]</td>
<td>u</td>
<td>u:</td>
<td>ur</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>[low]</td>
<td>e</td>
<td>e:</td>
<td>er</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>æ</td>
<td>æ:</td>
<td>ær</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>(V)</td>
<td>(V:)</td>
<td>(Vr)</td>
<td>(Vw)</td>
<td>(Vy)</td>
<td></td>
</tr>
</tbody>
</table>

of a glide as a privative feature), is 52, for representing all of the 21 stressed vowels in Table 3.13. At the same time, the degree of abstractness is not unreasonable.

The proposal also has a number of substantive virtues. First, it makes it easy to state the traditional short-long distributional dichotomy: short vowels lack a glide, long vowels have one.53 This, in conjunction with the dictum that stressed rhymes branch, explains the well known distributional fact discussed earlier: stressed short vowels occur only in checked syllables.

Like any set of vowel classes, the present set allows the comparison of English dialects in terms of mergers and splits of vowel classes in the set; unlike proposals in which the differences are expressed as different phonemes, the phonetic comparisons between corresponding classes are here usually (though not always) specified by rules that specify redundant phonological and phonetic details for an underlying phonological structure that is relatively uniform across dialects.

A problem with many abstract, pan-dialectal systems is that they are static, and do not contain within them the seeds of their own change. This is not the case for the phonological structure of Reference American, since many mergers, shifts, etc., are made sense of, phonologically, in this system. Thus /aw/-fronting in the South makes sense as a movement into a gap in the Vw system. The merger of Mary, merry simplifies a difficult-to-represent distinction. The raising of /æ/ is seen as initially due to a phonological shift of this historically short vowel into the long vowel slot emptied by the loss of the broad A (PALM) distinction. Most importantly, when coda consonants undergo phonetic change to become vocalic, the principle that only one glide can occur per syllable predicts

53In most English dialects "glide" is also a phonetically appropriate term for long vowels, since they do generally have phonetic glides.
severe restrictions on the possible phonological outcomes when the segment is analysed as a phonological glide. It is important to note that while these structural changes can be accounted for as changes in the abstract phonological pattern, there are also a great many low-level sound shifts in which the rules of phonetic implementation show social and historical variation. The system of abstract categories is one important level at which sound change occurs, but it is not the only one, as we will see in the phonetic patterns shown in later chapters.

The similarity of different dialects in the phonological structures presented here helps to explain the mutual intelligibility of English dialects. The problem of perception of other dialects has at least two facets. First, when a phonological comprehension error is noticed, the listener must reconstruct the intended sound. How this cognitive procedure works, and to what extent it does work, is an important and interesting topic of study, partly examined in the Cross-Dialectal Comprehension studies of Labov, et al. Second, in learning to understand a different but related dialect, listeners must be able to relate the sounds of the different dialect to the phonological categories of their own dialect. Listeners may map their own phonological structure onto a new phonetic space. How these processes proceed, and what are their limitations, are questions that go beyond the scope of this thesis; nonetheless, it is important to point out that these are important issues in studying the structure of phonologically related dialects.

The proposal accommodates both sides of the issue of whether to represent long vowels as /Vː:/ or /VV/. A vowel with an unspecified glide position may reasonably be written as /Vː:/, while the derived representation of the same vowel, after the glide has inherited vowel-quality features from the nucleus, may reasonably be written as /VV/. We have therefore removed from the horns of a false dilemma.

For convenience of reference and transcription, there is a simple symbolic representation of each of the vowels. Weird symbols are reduced to “a” symbols: ɔ, a, æ. Since the length-versus-glide argument is decided in favor of both (length = presence of a glide), people who prefer one or the other of /iː, eː, oː/ and /ii, ee, oo, uu/ and /iy, ey, ow, uw/ as convenient representations for the same vowel classes, can remain justified in their preferences, which amount to decisions about which level to transcribe in symbols. This entirely extra-theoretic, arbitrary decision can of course be made in any way one wishes.
to. I will occasionally use /iy, ey, ow, uw/ for /iː, eː, oː, uː/; ambiguity does not result.

This proposal provides a uniform surface phonological structure (the output of the lexical phonology, in the sense of Kiparsky 1982) which can be applied to dialects of English. Differences of inventory are to be stated in terms of mergers and splits relative to this inventory. Phonetic differences between the same categories across dialects can be stated in terms of the phonological rules that fill in these relatively underspecified feature-structures. It has considerable formal simplicity and elegance; it explains a number of interesting substantive patterns; it is not so opaque that it cannot be used for other purposes than construction of phonological theory. I will assume this phonological framework as a basis for the phonological and phonetic analyses of the remainder of this thesis.

3.6 Phonetic Implementation of Reference American

Phonetic implementation rules are an important part of the grammatical system. They are part of linguistic performance, but the competence/performance distinction is not very important at this level of the linguistic system. The mapping from phonological specifications to distributions of F1-F2 values in actual speech is undoubtedly very complex. In theories such as Lindblom’s, phonetic forms are a balance between maximizing auditory distinctness, which is a very complicated, non-linear, context-sensitive system of constraints, and maximizing articulatory ease, which is another complicated, non-linear, context-sensitive system of constraints. Developing an explanation for the patterns of surface distributions of vowel nuclei, patterns which result from the joint optimization of both of these systems, may be beyond our capabilities, since the workings of this optimization are only partly known.

Part of the research program of this thesis is to attempt to be clear about what these distributions themselves are. As an effective discovery procedure, one needs an approach that leads to interesting regularities. Without attempting to explain the entire system,

54In phonological and phonetic performance, errors are almost non-existent (Labov 1966), and dysfluencies or “false starts” themselves are well-formed according to a simple set of rules (Hindle 1983). Limitations of memory pose no difficulty for native speakers in the production or perception of sequences of sounds. Thus the flaws of performance data ascribed for syntactic data do not apply with any force to phonological and phonetic performance. On the contrary, in phonetics and post-lexical, surface phonology, a similar concern about errors in the data leads to the opposite conclusions: speech errors are more frequent when speech is self-conscious.
it is important to describe the surface distributions and to try to find what the surface patterns are. This approach could well be wrong, but there appear to be symmetries and patterns, and describing those patterns explicitly will lead to the regularities that do exist, in addition to a clear descriptive picture of the surface patterns. Some may object that we can’t expect to find anything, and they could be right. But if we don’t look, we won’t know.

In this section I discuss a mapping from the minimally specified post-lexical phonological structure to a more fully specified, default phonetic representation for this hypothetical dialect, which will be used for comparison with the phonetic forms in the particular dialects studied in later chapters. The phonological specifications of the vowels in Reference American are quite underspecified. This allows many of the differences between dialects to be stated in terms of differing rules for filling in the phonetic details from the cross-dialectally similar, underspecified post-lexical forms given here. Much work remains to describe the phonetic forms correctly, and to characterize the system that generates them.

Numerical phonetic implementation rules come in two theoretical flavors. First is the n-ary phonological rule, where categorical phonological rules are considered to operate on numbered categories (as in Labov, Yaeger, & Steiner 1972: 167ff). A quite different kind of rule is the kind of phonetic implementation rules found in Liberman and Pierrehumbert (1984), where phonologically specified targets are mapped onto F0 values according to general parameters such as pitch range, and the F0 contour is interpolated between the targets. The latter class of rules assumes some speaker-specific bounds on a continuous phonetic space, with continuous mathematical relationships (rather than discrete relationships among small-n sets of categories) describing the mapping of phonological structures into acoustic dimensions (which not coincidentally are derived from articulatory configurations that are continuously variable and subject to conscious control). The phonetic implementation rules to be described here are of the latter type.

I propose the following set of simple and rather general phonetic rules for specifying the phonetic forms of the vowels of Reference American. The proposal assumes that phonetic vowel space is a triangle (reflecting the acoustic facts), and that an important step in specifying the grammar of phonetic form is the specification of target vowel qualities for phonological categories in this phonetic space. Further steps involve phonetic timing.
and duration patterns, the effects of phrasal stress, and the coarticulatory effects of nearby consonants. Beginning with the features specifications in Table 3.13, then, the rules in Figure 3.4 apply. English-specific phonological rules operating on discrete abstract categories are marked with an asterisk (*), while the rest are proposed as universal phonetic implementation rules.

Figure 3.4: Phonological Rules and Phonetic Implementation Rules

(1*) Mark [low] glides as redundantly [rhotic].
(2*) Mark the redundant feature [round/back] as in Table 3.11.
(3) Link the quality features of nuclei to the unmarked glide timing slots.
(4) Locate those nuclei which are unmarked in the front-back/round dimension on the central line in the vowel triangle.
(5) Locate the [front] nuclei evenly spaced along the front edge.
(6) Locate the back/round nuclei evenly spaced along the back edge.
(7) Interpret non-low glides as phonetically raising glides, relative to the adjacent nucleus.

This set of phonetic rules generates the geometrical configuration in Figure 3.5, which corresponds closely to the phonetic qualities in Table 3.14. The rules need not be “linguistic”, in the sense of being language-particular; however, they do describe part of the system of phonetic implementation.

Table 3.14: Impressionistic forms of vowels as derived from rules (1)-(7).

<table>
<thead>
<tr>
<th>V</th>
<th>V:</th>
<th>Vr</th>
<th>V_y</th>
<th>V_w</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
<td>i'</td>
<td>u^a</td>
<td>i_r</td>
</tr>
<tr>
<td>e</td>
<td>e^i</td>
<td>o^*</td>
<td>e_r</td>
<td>o_r</td>
</tr>
<tr>
<td>ae</td>
<td>a</td>
<td>o</td>
<td>ae</td>
<td>a_r</td>
</tr>
</tbody>
</table>

This system generates phonetic forms from the post-lexical phonological specifications that are the input to the phonetic implementation system. This system, or one functionally like it, is implicitly assumed to map the phonological structure to phonetic performance. The reason for proposing such a system explicitly is the usual reason for formalism in linguistic theory: It is easier to see where the system is wrong, when it is made more explicit. Thus this rule-system is proposed in order that further linguistic and phonetic
research may build on it, by developing arguments for or against particular features of the system. This proposal represents progress in at least this sense.

Note that a phonetic form for /a:/ is not listed in Table 3.4, because I am not familiar with the phonetic forms for this vowel class in dialects that distinguish it. The remaining vowel qualities given are reasonably close to observable vowel qualities in some dialects.

This proposal provides an illustration of the division of labor between systems of phonetic implementation and phonological structure. Phonetic implementation is an independently necessary part of the system of speech performance. As argued at various points in this thesis, it also must be partly linguistic, because dialects differ in some of the details of phonetic implementation, and therefore speakers who acquire the speech forms of particular dialects must learn these differences as part of their knowledge of the dialect. Since this system is independently required, and partly linguistic, it makes sense to attribute some of the descriptive load to the structure of this system, thereby simplifying the phonological structure. One of the main lessons of the systems of phonetic implementation of pitch contours argued for in Liberman & Pierrchumbert (1984) is that n-ary pitch levels may be derived from an underlying binary phonological contrast. In this system, the front-back dimension is underlyingly represented by a single privative feature which may be present.
or absent. But on the surface there are seven distinct degrees of frontness, derived through
the system of phonetic implementation.

The output phonetic forms of this implementation system in Figure 3.5 and Table 3.14
form the reference phonetic forms which I will use as a baseline when describing vowels’
phonetic realizations in particular dialects as having “shifted” in one direction or another.

An important difference between the phonetic forms of Table 3.14 and those that are
often considered as the basic realizations of these vowels in “General American” has to do
with the long, non-low vowels. In Reference American as defined here, and furthermore in
most other English dialects, the non-low long vowels /i:, e:, o:, u:/ are phonetic diphthongs.
As pointed out to me by William Labov, this contradicts Lehiste & Peterson’s (1961)
declaration that the long high vowels in “American English” are monophthongs. “Within
the rather homogeneous dialect analysed, ... the syllabic sounds of American English can
be described in terms of fifteen syllable nuclei, subdivided into short and long. ...[The]
simple long nuclei are tense monophthongs...[i]...[u].” Examination of spectrograms given
in their article shows that there are in-glides for the realization of /i:/ in bead, bean, but
this is so far from the expected forms that they were not discussed. It turns out that their
subjects were speakers of a conservative white dialect in the Chicago area. In Wisconsin
and Minnesota, to the north and west of Chicago, not just /i:/ and /u:/ but also /e:/ and
/o:/ are in fact commonly monophthongs and sometimes even inglides. In Chicago, /ii:, u:/
are commonly monophthongs. However in most dialects of English, all of these long vowels
are up-gliding diphthongs. Lehiste & Peterson’s work is classic and often-reprinted, but it
suffers from narrowness of dialectal coverage. Phonetic facts from single particular dialects
should not be overgeneralized and attributed to the imagined homogeneous national dialect
of “General American.” The more general American facts are therefore represented in this
Reference American system, with phonetic glides on the high-front and high-back vowels
(depending on whether the vowel is front or back), as well as on the long mid vowels.

Notice that the glides are defined in relative rather than absolute phonetic terms. Thus
a [high] glide will glide from the nucleus to a phonetically higher position, but it will not
glide to phonetic high position unless the nucleus is at least mid. The [high] glides of [a̱c,
αː, o̱, i̱r, u̱, e̱, o̱]55 are each a step above their nuclei. Similarly when r-glides occur on

55 Here [y, w] are as Chomsky & Halle (1968) define them: higher than [i, u].
vowels, the r-glide itself retains the color of the nucleus: \([\text{i}^3]\) in \([\text{i}^3]\) is higher and fronter than the \([\text{a}^3]\) in \([\text{a}^3]\).

It is important to note that the nuclei of these long vowels are generally more peripheral and higher than the nuclei of their lax counterparts /i, e, a, u/, \(^{56}\) though this is not represented in the rules above, since these relationships are dialect-specific.

As Kenyon and Knott (1953) and later scholars have pointed out (e.g. Halle & Mohanan 1985) in various terms, vowel-vowel sequences (as for example N+G) often result in a relatively raised, tense, peripheral quality on the first vowel (the nucleus). This phonetic tendency, which may be called V/\_V Raising, applies to non-low Vr# sequences (beer, bear, boor, bore) but not 'VrV sequences within a word (syrup, berry, Barry, jury, sorry). It also commonly applies to the nuclei of the long vowels, consisting of NG sequences: /i:, e:, a:, u:/ all have raised, peripheral nuclei relative to /i, e, a, u/ (not shown in Table 3.14). This is the synchronic form of the historical sound-change generalization in Labov, Yaeger and Steiner (1972), that in chain shifts, peripheral (usually, long) vowels rise. A formal statement of the rule, which appears to apply under complex sets of conditions, and applies differently in different dialects and languages, is beyond the scope of the present work. Its effects, however, are pointed out in several of the discussions of the rules of phonetic implementation in later chapters.

V/\_V Raising also may be understood as the driving force behind the widespread sound change which raises lengthened /æ/ in many American dialects. If /æ/ becomes a long vowel (after /a:/ merges with other classes, presumably, since /a:/ occupies the long, low, front slot in Reference American), then it fits the structural description of the V/\_V raising rule, and the phonetic raising that operates on the nucleus of this vowel may be seen as deriving from V/\_V Raising.

Harris gives /æ/-tensing rules for Belfast, Philadelphia, New York City, considered to be deep, lexical phonological rules. Features like [+tense] cannot be added to lexical representations in lexical phonology because of structure preservation. The present analysis suggests that the tensing rule is instead a lengthening rule, with the peripheral and raised quality of "tensed" forms being analysed as phonetic features added later by the V/\_V

\(^{56}\)Central /a/ is not only quite non-peripheral, it also doesn't share the roundness feature with its counterpart /o:/, as in Table 3.11.
Raising rule.

Also, the tensing and raising of /ɪ, ɛ, u/ in the "Southern shift" (Labov, Yaeger, & Steiner 1972) can also be attributed to V-V Raising, if these vowels are treated as phonetically lengthened.

Actual systems of phonetic implementation are quite complex and subject to numerous influences, of stress, consonant context, etc. The rule-system provided here does not incorporate these influences, though it may be extended in order to do so.

The phonetic implementation rules given here are proposed as universal rules: they are not dependent on the linguistic phonetic patterns and processes that occur in different dialects. That is, up to this point the rule system proposed is consistent with the assumption of a universal phonetic implementation system. It will be seen in the descriptions in later chapters that this assumption must be modified, in order to generate the phonetic forms characteristic of particular dialects. The description of the interactions with adjacent segments, and other effects within the phonetic implementation system, what I sometimes call the phonetic grammar, is largely a matter for future research, with some steps in this direction taken in the following chapters. The regularities and patterns found in surface phonetic forms may lead to a deeper understanding of the phonological and phonetic system. It is to be hoped that this section provides a useful starting-point for investigation of these patterns.

3.7 Summary

This chapter has defined the English vowel system, first by enumerating the lexical sets which are useful in comparing the sound classes of English dialects, and then by structural characterization of the temporal and static aspects of vowel phonology in a useful, hypothetical dialect, Reference American. The result is an extremely simple, underspecified, phonological structure making use of privative features:

\[
\begin{align*}
\text{Vowel} &\rightarrow \text{Nucleus (Glide)} \\
\text{Vowel licenses Height ([\text{high}], [low]), ([front])}.
\end{align*}
\]
Thus, English vowels consist sequentially of a nucleus, followed sometimes by a glide. [front] and the height features [high] and [low] are licensed within vowels. Redundant features [rhotic] and [round] are filled in, it was argued, by language-specific phonological rules. Finally a set of universal phonetic implementation rules was proposed which relates the minimally specified phonological categories to actual phonetic forms. The phonetic and phonological patterns of Reference American will be used as a basis for comparison for the English dialects studied in the remainder of the thesis.
Chapter 4

Theoretical Background Issues

This chapter discusses theoretical issues that form an important background for this thesis. The distinction between linguistic and general phonetics is drawn. The social, linguistic, and historical centrality of the vernacular form of language is described. Some important theoretical issues in acoustic phonetics are discussed, such as the relations between acoustic categories — like the useful concept of “acoustic vowel” — and deeper categories; the definitions of allophony and coarticulation; some criteria for delimiting the set of acoustic phonetic features; and the basis for inferring audible differences from acoustic differences. Finally an experiment is described that repeats the fundamental work of Lisker (1949), with different results, which is attributed to the basic difference between phonetics and phonology.

4.1 Linguistic Versus General Phonetics

4.1.1 Types of Phonetic Facts

We may categorize phonetic facts into four groups: random, individual, dialectal, and universal. Random (low-level) performance facts include effects of coughing, true speech errors, etc. Individual facts include vocal-tract size and shape (from which are derived the frequency scale of formant space for a given speaker), some voice quality features, etc. Language- or dialect- or style-specific facts include the degree of rounding on labial consonants (Sapir 1921:43), many details of phonetic vowel quality (Labov, Yaeger, &
Universal facts include, for example, the absence in language of voiced glottal stop phonemes, or the presence in all languages of voiced sounds,¹ etc.

“Linguistic” is a word with two contrary connotations: “specific to language (in general),” and “specific to (particular) languages.” While universalists prefer the first connotation, the second is more useful in the context of phonetic description, where on the one hand, universal facts are most likely to be physical facts, not linguistic ones; and on the other hand, those facts that are characteristic of particular dialects or languages and not of others must be considered linguistic facts.

The linguistically relevant difference between speakers of one dialect and speakers of other dialects is that they have learned to speak the dialect in question. Dialect-specific phonetic patterns must be learned by speakers of the dialect in question. If speakers had not learned these patterns in some way, they would apply to all speakers, including those who have not learned to speak that particular dialect.

In short, those phonetic details that are specific to particular languages or dialects must be accounted for as part of linguistic phonetic descriptions. More universal principles of general phonetics may account for other phonetic details, which need not be specified in the grammar of a particular language. Phonological (or linguistic-phonetic) theory and description can stop only where general phonetics begins — that is, well within the purview of contemporary phonetics, which studies both general and language- or dialect-specific phenomena.

On the other hand, there may be intricate interactions between universal and dialect-specific principles in the phonetic pattern of a language, so that a full understanding of its surface-observable phonetic characteristics would require simultaneous study of both kinds of principles.

### 4.1.2 The Overlap of Phonetics and Linguistics

Single phenomena are both phonological and phonetic; linguistic phonetic effects like dialect-specific coarticulations and vowel quality differences are documented in this thesis. There may in fact be a boundary between linguistic and general facts about the sounds of language. This boundary may be called the boundary between linguistic phonetics and

¹This may be due to the fact that sounds are generally louder when voiced.
general phonetics. This is certainly not the boundary between phonology and phonetics, as commonly understood; it is at a much lower level. Contemporary phonetics straddles the boundary, including both linguistic (language-specific) and more general phenomena. How much of phonetics is contained within linguistics depends on the extent to which phonetic detail is language- or dialect-specific. The details of the location of this boundary within phonetics are empirical matters. But there is no doubt that a useful approach to understanding the limits of linguistics is to learn as much as possible about all aspects of phonetics, whether they turn out in the end to be properly linguistic or non-linguistic (general) facts. Taking the opposite approach — excluding all of phonetics from linguistics just because some of phonetics is non-linguistic — is fruitless.

4.2 The Vernacular

This thesis focuses on phonetic phenomena in fluent vernacular speech. For a definition of the vernacular, see Labov (1966). In short, it is an un-monitored speech style, a style in which the speaker pays attention to the content rather than to phonological or grammatical form. Thus the speaker is caught up in what he or she is saying, and pays little attention to how he or she is saying it. The differences between this and more formal styles of speech in some segments of a community are taken to an extreme in the speech of certain social groups, which are in this sense the most vernacular speakers. For example, in New York City, increased r-lessness is associated with informal styles for individual speakers, and also occurs at the highest rate among working-class speakers. Vernacular /r/ behavior in New York is both a pattern of style-shifting and a pattern of social variation (Labov 1966).

Why study vernacular, fluent speech? What are the differences between the vernacular and “laboratory speech”? Laboratory studies frequently use fully-stressed citation forms, often single-word utterances making up a complete intonational phrase; they have often been rehearsed by the speaker well in advance of the production; they may be nonsense words; they are typically read from a prepared text or memorized. Yet these factors may obliterare regular phonetic conditioning, as shown by Keating and Huffman (1984) and Labov (1986). The effect of self-reflection on one's speech is to introduce unsystematic "correction" of the phonetic forms produced — which amounts to bias in the data. On the
other hand, the regularity of phonological patterning in vernacular speech as opposed to formal, or highly self-monitored speech, has been persuasively argued for (cf. Labov 1966, 1986). Further, fluent vernacular speech is language's most basic, systematic, as well as frequent form. The vernacular is fundamental to the history of languages, since it is the basis of their historical continuity. The vernacular is fundamental to the social meaning of language forms, since speakers are (socially) defined, generally for life, by the vernacular dialect acquired in their youth.

Eckert (1989) and others show that sound change takes place in the young (adolescent) peer group. In the second decade of life, the phonetic forms that speakers will continue to produce regularly throughout their lives are molded through the heat of peer pressure (or perhaps by the burning desire to conform), a process which also gives sounds their most deeply-felt social meanings. It is the local vernacular dialect which is acquired in this period, and which is itself gradually modified by both social and linguistic forces of change.

Thus, the vernacular local dialect is the locus of sound change. It might be argued that only a dialect that has historical continuity is "real," or historically significant. Consider a somewhat artificial example: suppose (incorrectly) that (what is normally thought of as) Latin and French are directly related as ancestor to descendant, i.e., that they were historical stages of a single, continuous, living language. The synchronic and historical studies of phonological variation and change by Labov and his students and colleagues may be extended to establish the Latin/French vernacular as the site of the changes that differentiate these historical stages. In this case, the true French and the true Latin, which are related concretely, historically, are vernacular French and vernacular (spoken, "Vulgar") Latin. Book-Latin and book-French are not related in this direct, concrete way, and are therefore mere historical sideshows, dead remnants of the historically continuous, vernacular, living language.

Thus it is the vernacular that is the historically significant, or "real" form of language. This a priori historical argument has added merit if we consider the associated facts that most people are vernacular speakers, that local prestige of individuals is correlated with their degree of advancement in use of the changing vernacular (so that the local people that
are most important, in some sense, are the ones most advanced in their speech)\(^2\) and finally, that the most regular and systematic form of language is the vernacular (Labov 1966). For these reasons vernacular speech is an important object of linguistic study. The core of this thesis consists of phonetic results derived from the frequent, natural, systematically conditioned, socially meaningful sounds found in everyday, fluent, vernacular speech.

4.3 Issues in Acoustic Phonetic Research

4.3.1 Definitions

This section discusses and defines the observationally useful acoustic segments, *acoustic vowel* and *acoustic consonant*, as well as certain other phonetic terms which it may be helpful to clarify.

Acoustic segments which can be easily be identified and delimited on spectrograms are analytically convenient objects, but their status in a theory of linguistic performance is far from clear. It is quite unlikely, for example, that the timing patterns of acoustically well-defined events constitute, in themselves, the underlying speech timing pattern. Phonetic timing data are observationally convenient, and form the raw materials for studies of linguistic timing. But analytically convenient fictions do not necessarily reflect the true categories of the phonological production mechanism. Many patterns will be artifacts. For example, the relation between underlying speech timing and the acoustically well-defined moments like burst-onset or the moment of voicing onset, etc., is unclear, as is evident from continuing difficulties in establishing phonetic reality for "syllable-timing" and "stress-timing" patterns in various languages. Similarly, the timing of "magic moments" like burst onset, voicing onset, and other consistently measurable parts of the speech signal, must be related to or somehow derived from intonational structure, inherent segmental timing patterns, etc., in a temporal theory of phonetic implementation, but such sharply-defined acoustical segments are not themselves the fundamental underlying units of production.

Despite these reservations, acoustic segments are the basic data for much speech research, and in particular for this thesis. We must be clear about what we are talking about; therefore definitions are important.

\(^2\)cf. the Project on Linguistic Change and Variation, Labov (1980).
It is also clear that acoustic segments are not unrelated to the underlying production mechanisms. Burst onset, for example, occurs at a particular point during the tongue- or lip-gesture that closes and opens the vocal tract at a particular place of articulation. The burst is distinct from the gesture, but it may be used to locate a part of the gesture in time, and the spectral and temporal information in the burst carries perceptual cues as to the underlying gesture that occurred. Further, there are exceptions: those gestures which typically have particular acoustic events associated with them don't always co-occur with these events. For example, expiratory force may be so weak in some performances that the burst is unmeasurable. Thus, typical patterns may not accurately describe all instances. Nonetheless, we proceed by examining and characterizing what we think are typical instances, pointing out along the way those cases that are excluded or that do not fit the typical patterns. The existence of exceptions does not prevent us from characterizing the typical patterns, and from basing a partial understanding of the underlying system on these patterns. This is the strategy in this thesis, which is primarily a study of acoustic vowels.

- An **acoustic vowel** is an acoustic segment where the noise source is voicing, and simultaneously where there is neither oral closure nor sufficient constriction to produce turbulent airflow.

- An **acoustic consonant** is an acoustic segment that corresponds to oral closure, frication, burst, or aspiration.

- The **realization** of a phonological element: Any acoustic feature which provides a perceptual cue to an underlying segment partly constitutes the realization of that underlying segment.

- **Coarticulation** is generally used to refer to any phonetic characteristics which are simultaneously determined by more than one underlying unit.

The reason acoustic vowels are interesting is that we understand fairly well how they are produced, and we can infer from their acoustic structure what the mouth is doing, in some detail (cf. Chapter 2, Acoustics). On the other hand, during noise-excited segments, the formant structure is typically less clear (F1 during [s], for example, is hard to locate
precisely by any objective means, as a spectrogram will make clear), and thus articulatory movements and audible vowel color are less clear.

Therefore acoustic vowels are a very useful object of description. However, as follows from these definitions, the realization of an underlying vowel is not composed entirely of the corresponding acoustic vowel; nor is the realization of an underlying consonant entirely constituted by a corresponding acoustic consonant. Coarticulation occurs very widely in the realization of underlying sounds. Consequently, acoustic vowels frequently contain some aspects of the realization of underlying consonants, just as acoustic consonants often contain aspects of the realization of underlying vowels.

Acoustic vowels are bounded by the beginning and ending of voicing, by onset or offset of frication or oral closure. An acoustic vowel between stops typically begins at the point of post-release voice onset and ends at the point of devoicing or oral closure. Not all vowels are associated with acoustic vowels. Whispered vowels and [h] are not acoustic vowels, nor are devoiced vowels, which most frequently occur (in English as well as in Japanese) between voiceless obstruents, when the vowel is high, unstressed, and front. Such vowels are indeed (underlyingly) vowels, but their formants are difficult to measure, and the movement of the articulators is harder to follow by formant tracking. Because of these difficulties, such vowels are excluded from acoustical study here.

Semivowels are a problem for acoustic segmentation. They are often difficult to delimit in time. Where there are sharp acoustic discontinuities at the boundaries of the semivowel, as with the release of apical [l], those discontinuities define the border between the acoustic semivowel and an adjacent segment such as a vowel. Dark [l] without apical contact, and thus without clear acoustic boundaries, is an acoustic vowel, while /l/ with apical contact, and sharp acoustic discontinuities at the point of closure is not. Where there is a definite non-syllabic steady-state segment (as in /wɛ/ where the /w/ may be held in a steady state for some time), the onset of change out of the steady-state into the vowel is the location of the beginning of the acoustic vowel. Postvocalic English /r/ is an acoustic vowel.

It may seem plausible to distinguish between linguistic allophony, in which formal phonological rules of arbitrary expressive power modify abstract phonological segments (symbolic feature structures), and phonetic coarticulation, in which phonetic segments
(articulatory gestures) influence each other by virtue of the physical limitations (e.g., sluggishness) of the physical speech production system. However, a middle view is in fact more sensible, in which phonological rules require limitations on expressive power that reflect physical limitations of the speech production device, and in which the “physical limitations” of the device may vary from one language or dialect to another. In such a view, a coarticulated phonetic segment is just an assimilated allophone. “Coarticulation” and “segmentally conditioned allophony” are in this view identical. Perhaps the most important distinction between the terms is not in the phenomena referred to, but in that “coarticulation” is more common among phoneticians, while “allophone” is more common among phonologists.

The adjacent underlying consonants AND the underlying vowel simultaneously determine the changing formant structure at the edges (as well as in the middle) of acoustic vowels. These transitions are part of the realization of the underlying consonant as well as of the underlying vowel, as recent perceptual studies have amply demonstrated (Strange, et al 1976, and references in Strange 1989). A transition is an important cue to both the underlying consonant and vowel of which it is a joint realization.

Defining “acoustic consonant” and “acoustic vowel” allows us to say that coarticulatory effects on an acoustic vowel are effects of an underlying (articulatory or phonological) consonant, without implying that the “actual” consonant and vowel are realized elsewhere. An acoustic vowel segment is not to be taken as corresponding exclusively to an underlying vowel; it may, and here does, correspond to (i.e., provide perceptual cues for) underlying consonants. Thus coarticulatory effects observable in acoustic vowels are not incidental effects of a consonant on a vowel, but are part of the phonetic realization of the underlying consonant itself.

What is “easy” varies across languages. Just as what is physically easy for some is difficult for others, because of the practice to which their muscles — or neuromuscular control systems — are accustomed, different languages and dialects may set the default energy-expenditure level of the various speech articulators at different levels. In this way, the physical sluggishness of the articulators may effectively vary across languages. Cf. Sievers (1901), who offered the explanation of phonological symmetry that there is a different rest position of the organs in speakers of different languages. We may add to Sievers that certain muscle movement patterns are highly practiced in one language and not in another, and are therefore easier, both because of the resulting strength and endurance of the muscles involved, and the more redundant, robust, and fully-trained neural control systems behind the movements involved.
4.3.2 Acoustic Phonetic Features

The linguistic and psychological significance of acoustic segments like those defined above must be explored empirically. The reason for supplying these definitions is that acoustic studies commonly investigate the properties of such segments, and it clarifies matters to define what we're talking about. This thesis, in particular, is about the properties of acoustic vowels. There is certainly information about vowel identity beyond the temporal limits of the acoustic vowel; whispered vowels can be identified, yet they do not constitute acoustic vowels; similarly, one can often identify vowels from the spectral information contained in acoustic consonant segments (for example, F2 is often clearly identifiable during adjacent turbulent noise segments). However, in phonetics, unlike in phonology, information is redundantly present. If a vowel's identity is signalled in many different ways, then a listener who is sensitive to all the redundant cues will be more easily and confidently able to identify it. Indeed, if some cues are effectively eliminated through environmental noise, listener's poor hearing, etc., then it is only the redundancy in the signal that enables the sounds to be correctly identified and for communication to function robustly.

The fact that information is contained in very short transitions into and out of a vowel does not imply that the "nucleus" of the acoustic vowel contributes nothing to perception. Similarly, the absence of an acoustic vowel corresponding to a phonological vowel in an utterance does not imply that there are no phonetic cues for listeners to make use of in identifying that phonological vowel. Many disparate sources of information about the underlying linguistic forms may be used to understand spoken language. The features may be "distinctive" or not, but as long as they are useful to listeners in constructing the linguistic form of an utterance, they are real, phonetic features.

This argument can be taken too far. It remains important to ask, What are the true cues? It is well and good to say that many redundant cues are present, but constructing the psychologically correct inventory of cues is another matter. Any phonetic measurement that can be made constitutes a potential member of this inventory, and many are clearly false members. For example, the formant frequencies during the thirteenth pitch period after voice-onset in a stop-vowel sequence constitute a potential cue for vowel identification.
Indeed that information may often be enough to identify vowels correctly, or certainly to reduce the set of possibilities. However, depending on rate, pitch, and other factors, there may be fewer than thirteen pitch pulses in the realization of a given vowel; or the thirteenth pitch pulse may be still in the onset transition. An infinity of such possible measurements are derivable from the acoustic vowel's formant trajectories and duration (e.g., the first pitch period, the second, the third, . . . ). The catalog of cues may be analytically reduced to a smaller set from which the larger set is derivable, and this reduction is an important part of the progress of speech science. There is a balance between a senseless proliferation of cues and an overly abstract view of speech perception and production: the catalog should include all the cues that make a perceptible difference, and exclude all measurements that can be derived from others.

The question then arises whether patterns of formant frequencies during acoustic vowels constitute a legitimate part of the ultimate space of phonetic dimensions. If the mechanism of speech perception includes an F2-tracker, for example, it is possible that it is insensitive to rapid fluctuations in amplitude and to the noise source, and thus that it doesn't stop tracking F2 at the exact moment that voicing ceases (i.e., at the edges of acoustic vowels). Thus segmentation points that are acoustically well-defined may be irrelevant at various levels of the speech perception process. Similarly in production, the ongoing gestures of tongue, jaw, and lips that give rise to the acoustic realization of a vowel may partly occur during unvoiced segments outside the acoustic vowel, because the glottal voicing gesture is absent. The articulatory gesture for a vowel may start and stop at points that are only indirectly related to the temporal bounds of the acoustic vowel. Thus the particular properties of acoustic vowels chosen for measurement and presentation in this thesis are not necessarily the ultimately correct cues.

However, in the process of learning about how speech production and perception work, we must make progress where we can, by investigating promising sources of information, and by exploiting the knowledge that we do have, in order to learn more. The patterns uncovered in an exploration of formant structure in acoustic vowels must be determined in some way by phonetic grammar. The issue here is not necessarily to resolve immediately the ultimate psychological questions about the true catalog of phonetic cues for speech perception, but to find interesting patterns in measurements that call out for explanation,
to model them (by writing programs or rule-systems that generate similar patterns, for example) and ultimately to predict unobserved patterns from the understanding gained.

4.4 Phonetic Inference from Differences in Formant Frequencies Among Classes of Vowels

Two methods of characterizing phonetic vowel qualities are used in this thesis. Impressionistic transcription of vowels according to a standard set of vowel qualities such as those of Daniel Jones is very useful, but only when readers are thoroughly trained to recognize the precise auditory qualities of the symbols used. Unfortunately, this means that phonetic descriptions are unreadable except to a relatively small, properly trained audience. Efforts to expand this audience by making it easy to learn standard vowels include Jones' original publication of a 78rpm record of the Cardinal Vowels, and my Macintosh program, "Jones' Phones", which trains the user in identification of these vowels. A limited amount of IPA impressionistic transcription is presented in this thesis, since it is useful for calling to mind the phonetic qualities without referring to charts (though it is only interpretable when the reader knows the values of the symbols).

Another time-honored method of characterizing vowel quality is through plots of measurements of the first and second formant frequencies: the F1-F2 chart. Formants are interesting for several reasons. They are more objective than impressionistic categories, especially when the sounds measured are relatively short in duration. They are as good as any other spectral representation (under the assumptions of an all-pole model). They distinguish audible vowel qualities extremely well, as shown in the next section. And they can be used to infer articulatory configurations, as justified in detail in the Acoustics chapter. Thus they are good representations for static vowel qualities, analytically, perceptually, and articulatorily. Therefore charts of F1 vs. F2 are the core of the data presented in this thesis.

In 1949, Leigh Lisker wrote perhaps the first phonetics dissertation to make use of the

---

4In an all-pole model, the representation of a spectrum in terms of center frequency and bandwidth of spectral peaks, or poles, is interconvertible with several other representations. Thus formants may be considered as good as any other analytical representation of the speech spectrum, under the assumptions of this model.
recently declassified spectrograph machine. He made the disconcerting discovery that the main cues to vowel quality, namely the frequencies of the first two formants, were unable to distinguish reliably between clearly distinct pronunciations of two vowel phonemes. The phonemes involved are adjacent in formant space — the vowels in the words *pap* and *pep*. Since he made sure each token was clearly identifiable as the intended form, throwing out tokens which were ambiguous to the ear, the similarities between */æ/ and */ɛ/* were not so great as to make any token of one sound like a token of the other. This result is one of the fundamental facts about F1-F2 charts: vowels that sound different may overlap in F1-F2 space ("formant space"). How then can we trust formant measurements as reliable representations of vowel quality?

Lisker (1949) prevents us from making a certain class of inference, that vowels which overlap in formant space are in fact phonetically identical, since */æ/ and */ɛ/*, for example, may overlap but sound quite different. Similarly, */ɛ/ and */ɪ/* may overlap in F1-F2 space: */ɛ/* is a long up-gliding vowel, while */ɪ/* is a short, frequently ingliding vowel. They may pass through the same location in formant space, but they are going in opposite directions. Thus the fact that measurements at single time-slices show overlap between the two categories is not evidence that the sounds are phonetically identical.

The purpose of the next two sections is to argue for the validity of the converse inference: vowels occupying significantly different parts of formant space cannot be phonetically identical.

4.4.1 Mismatch between Audible Differences and Differences in F1-F2

This thesis makes use of a particular kind of inference, namely, that significant formant frequency differences reflect real differences in phonetic vowel quality. This section attempts to show that this kind of inference is valid. To do so, we must consider what are the relations between "same" and "different" in the realms of perceivable sound quality and of acoustic measurements, as displayed in Table 4.1.

Two classes of sounds may be the same on both grounds (Same), or different on both grounds (Different). In these cases, measured differences reflect auditory differences, and overlapping measurements reflect auditory identity. These constitute the ideal situations,
Table 4.1: Same and Different in Perception and Acoustics.

<table>
<thead>
<tr>
<th>measurable?</th>
<th>audible?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Different A</td>
</tr>
<tr>
<td>No</td>
<td>B Same</td>
</tr>
</tbody>
</table>

in which both ways of characterizing vowel quality are mutually consistent.

But there are also non-ideal cases. Two kinds of possible mismatch may occur between F1-F2 measurements and auditory quality. The measurements may be indistinguishable while the classes sound different (A), and the classes may sound the same while the measurements are different (B). In case A, we generally assume that other features of the sounds besides measurements of F1 and F2 at the nucleus allow them to be perceived as different. For example, measurements of the nuclei of *bait* and *bit* can be entirely overlapping, but one is longer and up-gliding, while the other is shorter and in-gliding. The result is that they sound completely different, although their nuclei may lie in the same part of F1-F2 space. Similarly, Lisker's *pap* and *pep* tokens may have differed in duration, gliding, etc.

Case B, in which the measurements are different but the sounds are perceived to be the same, comes in two flavors. In one flavor, the perceived sameness of the sounds is due to their phonological identity, not to their phonetic identity. When undergoing raising, /æ/ as in *pat* or *pan* commonly varies in natural speech in Northern U.S. dialects from [æʰɔ] to [ɛʰɔ], depending on consonantal context. Native speakers of these dialects have learned to classify these different sounds as the "same sound", i.e., as the same phonological unit, despite the significant phonetic differences between them. But despite this linguistic effect in perception, by which different sounds are perceived as identical, the sounds remain phonetically different. Children, trained phoneticians, adults listening to the sounds taken out of a linguistic context, and speakers of dialects where the phonetic change has risen to popular consciousness and become stigmatized, may differentiate them. Thus the first flavor of the mismatch between perception and measurement is one in which perception
is wrong: the sounds really are phonetically different, as the measurements show, but the linguistic system of the listener may inhibit the recognition of this fact. Here, the inference from “acoustically different” to “phonetically different” is valid, despite the protests of phonetically confused native speakers.

The second possible flavor of this mismatch is where the measurements are wrong: the sounds are measured as different, but they actually are auditorily indistinguishable. This situation can arise in two trivial ways, and apparently not otherwise. First, the measurements can simply be erroneous. If the measurement procedure leads to picking the wrong formant, for example, one vowel can appear as another. Thus if a nasal formant at 400 Hz is present as well as an F1 at 300 Hz, then [i], where \( (F1,F2) = (300,2200) \) can appear as [u], with \( (F1,F2) = (300,400) \), if these two resonances are chosen as F1 and F2. Such errors can be avoided by carefully matching the formants chosen with what is known about the relation of vowel quality and formant-frequency. This makes formant-tracking an art rather than a science, in which the phonetician must go back and forth between looking at the spectrogram and formant-tracks to listening to the sound itself. Cases of bad formant tracking must be identified auditorily.

The second trivial source of the second flavor of mismatch between perceived and measured vowel quality derives from the inherently limited degree of precision of measurements and of perception. If two measurements, expressed numerically to the nth degree, differ by less than the measurement error, then the small numerical difference between them is not significant. Also, the human perceptual system has a limited sensitivity to formant-frequency differences, not a great deal different from the measurement error. Differences below this “difference limen” (Flanagan 1955) are insignificant.

The last conceivable situation, in which this kind of mismatch could occur, is if the mismatch were genuine, that is, if the sounds really were auditorily identical despite non-erroneous, sufficiently large (greater than the difference limen), statistically significant differences in formant frequencies. I know of no good evidence that such cases occur. The most common class of evidence is that one can’t hear a difference that one can perfectly well measure and see. But this can be attributed to a lack of phonetic sensitivity of the individual listener, rather than to the lack of an actual phonetic difference. In fact, listeners with a low degree of phonetic sensitivity, who believe that two quite different
sounds (that may be phonologically the same) sound the same, can often be convinced by rapid interactive playback of the distinction. A method for learning to perceive subtle differences is to make an interactive vowel chart. In an interactive vowel chart, tokens are displayed on a computer screen in the form of labelled buttons that are located on an F1-F2 plot. When a button is activated by selecting it (with a mouse), the sound associated with the button is played out over a loudspeaker, so that rapidly going back and forth between two or more tokens, playing them out repeatedly, and listening for the differences in quality between them, will bring out the finest audible differences. Practice with interactive vowel charts can help develop listeners’ phonetic sensitivity.

True class-B mismatches do not seem to occur. For F1, F2 differences between two sound classes to be judged both significant and audible, the difference must be both statistically significant and greater than the threshold of audible differences (Flanagan 1955). When I find F1-F2 differences between sound-classes that meet these criteria, I will infer that they reflect genuine, audible phonetic differences.

4.5 Lisker (1949) Replicated

A replication of the experiment in Lisker’s thesis was carried out, which makes two points. First, the phonetic difference between two linguistically different sounds can be rather small. Second, pronunciations of the “same” sounds by speakers of different dialects can be phonetically different and thus lead to different results. The experiment may also be taken as an example of the improvements in speech-processing technology since the spectrograph machine.

This replication of Lisker (1949) used myself as the speaker, and more modern methods than were available then. 114 tokens of the form [pæp], and 142 tokens of the form [pcp] were spoken in isolation and a good quality cassette tape-recording was made. Each syllable was spoken as a complete utterance with its own separate intonation contour; a pause separated each adjacent pair of tokens. I took care to produce cardinal [æ] and [ɛ] qualities. I transcribe the pep tokens as [pʰɛp] and the pap tokens as ranging from [pʰa³p]

---

5The tape recorder was a high-quality (Nakamichi) cassette recorder; the tape used was Maxell XL-11 cassette tape; the microphone was a broadcast quality Shure 570S lavalier mike; the setting was an isolated room in a quiet house, with no machine noise.
to [pʰæːp], with most tokens as [pʰæp]. My intent was to produce the [æ] in *pop* not as many North Americans produce the vowel in *TRAP* — namely as a vowel more-or-less front-raised away from [æ] — but as a true [æ] sound. While I succeeded in producing auditorily low-front [æ] tokens, they vary somewhat in how low they are. The result of this is that F1 measurements of the [æ] tokens are more widely dispersed than the F1 measurements of the [ɛ] tokens, which I produced more naturally.

The tape-recording was digitized at 12,000 samples/second, with a resolution of 15 bits. The resulting waveform was subjected to automatic segmentation of the onset and offset of the acoustic vowel in the following way. First, the RMS amplitude value was computed 200 times per second using a window 20ms in duration. Then an amplitude threshold was set to an effective value, and the beginnings and endings of the onset of the acoustic vowel were located according to a “sloppy threshold-crossing” algorithm developed for the purpose.\(^6\) This method, with appropriate thresholds, is used to generate segmentation points that define the acoustic vowel. This method is quite accurate because of the abrupt discontinuity in the amplitude contour at the onset and offset of the acoustic vowel; an appropriate threshold can be determined by inspection of the contour at a few boundaries. After the algorithm was run, the results were examined for errors by hand; a slightly different threshold may be chosen, to improve the overall accuracy of segmentation. Finally, gross segmentation errors were eliminated by hand.

Next the formant trajectories are computed using the program, *formant*, and the default parameters described in step 4 of the 12-step program, page 136. Then the trajectories of the first three formant frequencies are extracted, within the segmentation points delimiting the acoustic vowel. These formant trajectories are then time-normalized, so that time t=0 occurs halfway between the endpoints. Finally, all tokens of each class are plotted on a time-frequency display, as shown in Figure 4.1. Each dot on these charts

---

\(^6\)Since a moving-average amplitude contour may be spiky rather than smooth, some method is necessary for separating “true” crossings of the threshold signaling the beginning of the acoustic vowel, from false crossings due to short-duration amplitude spikes due to single glottal pulses, transient bursts, etc. “Sloppy-crossing” is an algorithm developed for this purpose which requires the threshold to be crossed for more than a given fraction, \(f\), of a given amount of time, \(A\), before a “true” threshold crossing is considered to have occurred. If these parameters are set to \(f=80\%\) and \(A=.1\) seconds, for example, then if the parameter stays over the threshold for 80% of any .1-second segment, a threshold-crossing is recorded at the beginning of that segment.
represents a single estimated formant frequency at a single time-offset relative to the middle of a particular token. Thus a single vowel token of .10 seconds duration would have 3 formants * 10 frames of .01 seconds each = 30 measurements displayed on the chart, where the three formants for a given frame are displayed one above the other, and frames are displayed sequentially, .01 seconds apart on the time axis.

The two charts represent the distribution of formant trajectories within the syllable for the two forms pap and pep. First, observe that the distributions are quite consistent across tokens, especially in F1 for pep. There is a clear trajectory which all tokens follow quite closely.

Second, the differences between the pap trajectories and the pep trajectories seem at first glance quite small. Qualitatively this is certainly true: both have a similar-shaped rise in F1 and fall in F2 across the syllable; they even share a similar dropoff in F3 frequency near the end of the vowel. These movements reflect a slightly ingliding pronunciation, noted above as [ɔ].

Third, an apparent difference between the two forms is that the pap tokens seem considerably longer than the pep tokens (mean durations are ~ .17 seconds versus ~ .14 seconds). However this difference is mostly explained by other factors. The set of pap tokens was produced before the set of pep tokens; as time went by the production of tokens speeded up, from about .713 seconds per utterance for the pap's to about .600 seconds per token for the pep's. Thus 84% of the duration differences can be attributed to an increase in overall rate rather than a difference in intrinsic duration of the two vowels. That is, while a small part of the decrease in measured duration from /æ/ to /ɛ/ may be indeed be due to intrinsic duration differences, most of it is due to the simple fact that I was talking faster for /ɛ/ than for /æ/. Normalizing for rate, the difference between the two classes in terms of duration may be unreliable.

Fourth, despite these similarities, the two classes are completely distinguishable, as shown in Figure 4.2. When plotted together on the same chart, the pap and pep F1 trajectories have virtually no overlap. This chart shows that a clear difference in vowel quality can be reflected in quite small (about 110Hz), but genuine differences in formant frequency measurements.

Lisker's (1949) results differ from the current results. There, considerable overlap in
Figure 4.1: Overlaid formant trajectories for *pap* and *pep*, separately.

**pap**

![Graph of overlaid formant trajectories for pap](image)

**pep**

![Graph of overlaid formant trajectories for pep](image)
Figure 4.2: Overlaid formant trajectories for \textit{pap} and \textit{pep}, together.

\begin{center}
\textbf{pap and pep}
\end{center}

\begin{center}
\begin{tikzpicture}
\begin{axis}[
    xlabel={Time (seconds)},
    ylabel={Frequency (Hz)},
    xmin=-0.1, xmax=0.1,
    ymin=500, ymax=1500,
    xtick={-0.1, -0.05, 0, 0.05, 0.1},
    ytick={500, 1000, 1500},
]
\addplot[only marks, mark size=0.5pt] table [x=F, y=S] {data.csv};
\end{axis}
\end{tikzpicture}
\end{center}
F1-F2 space was found between measurements of pronunciations of these two words. Here, on the other hand, there is virtually no overlap in F1. One explanation for the differences could be the methods of measurement. Measurement by ruler and eye on a printed spectrogram includes an inherent randomness of ±25 to ±40 Hz even for practiced phoneticians working on good spectrograms with distinct formants. The difference found in these charts is not much greater than this inherent error, so perhaps the hand-measurement methods used were responsible for the overlap Lisker found. On the other hand, an equally likely explanation is that Lisker's pronunciations of *pap* and *pep* at the time were simply more similar to each other than mine were. This explanation drives home once again the difference between phonetics and phonology: *pap* pronounced by Lisker is the same word as *pap* pronounced by me; they have the same phonemes in them. But phonetically, my pronunciations and his may be quite different. Lisker is a Philadelphian; his /æ/ phoneme may therefore be raised from cardinal [æ], while I was making a conscious attempt to produce cardinal [æ]. So just because they are the same (lexically) doesn't mean they are really the same (phonetically). The same word, with the same phonemes, in the same context, can systematically be pronounced in different ways.

To summarize this replication of Lisker (1949), I find that while *pap* and *pep* in my pronunciations are almost identical in their formant trajectories (F2 and F3 are identical, and the F1 trajectory shape is identical, and quite close in frequency), nonetheless the F1 trajectories have almost no overlap at all, and differ by about 100 Hz in F1 alone. Thus a rather small but consistent difference in just a single formant frequency, can reflect a very clear difference in vowel quality.

The above experiments provide a valid instance of the inference argued for above: rather small differences are found to be extremely consistent, and reflect genuine, audible differences in vowel quality. Lisker (1949) makes the same point: two classes of sounds which are audibly quite different may overlap in F1-F2 space, but have different means. The bottom line, then, is that significant differences, even small ones, between measured formant frequencies between classes of vowels can be taken seriously as reflecting real differences in phonetic quality.

Formant frequencies, and the articulatory dimensions of mouth opening and tongue

---

body frontness, are continuously variable, unlike the discrete (usually binary) dimensions of phonological structure. It is controversial whether the control humans exert over these continuous dimensions in speech production is part of their linguistic competence or not. This hypothesis is implicit in the work of Labov, Yaeger and Steiner (1972) and in this thesis. Humans do demonstrate fine-grained, precise, conscious control over tongue movement in the production of sounds over continuous (or at least, fine-grained) dimensions of articulation and acoustics. The (non-universal) ability to whistle scales shows that that such control is indeed a part of human sound production capabilities. Thus this kind of control is certainly part of human phonetic competence. The question of whether it is also part of human linguistic competence, is explored in later chapters.

4.6 Vowel Reduction

A striking difference between vernacular and formal speech is the massive and systematic phonetic or low-level phonological reduction that occurs, deleting or modifying as much as 25-30% of the sounds expected on the basis of words' underlying forms (Veatch 1988). This commonly occurs with neither speaker nor hearer noticing it or often even being able to notice the deletions and reductions on having them pointed out. It is striking how different are the sounds we think we produce and hear from the sounds actually produced, which only careful listening for the phonetic reality will reveal. Prominent examples of reduction in dialects of English include flapping of /t/ and /d/, stops becoming fricatives, liquids becoming vocalic and other consonant manner lenitions; /r/-devoicing in voiceless clusters (as in “tree”, where /r/ may sound more like [ʃ] than a resonant retroflex [ʂ]); voicing assimilation of (word-final) obstructs (as in the plural /z/ in “the numbers seem” — cf. Veatch 1990); changes that take place when obstructs are juxtaposed through the deletion of intervening vowels; and vowel reduction, in quality and in amplitude, duration, pitch (range), and voicing. Perhaps the most technically tractable of this class of phonetic reduction phenomena, which may also display both dialect-specific and general effects, are the phenomena of vowel reduction, a central aspect of this thesis.
4.6.1 Types of Vowel Reduction

While vowel reduction may seem at first glance to be a single phenomenon, it is important to distinguish the many types that exist. Phonetic vowel reduction is not the same as phonological vowel reduction, and there are subtypes of each. While phonetic vowel reduction is considered to be a universal phonetic phenomenon, there are types of phonological vowel reduction that are quite different from language to language. We will discuss the types of phonological reduction in order to distinguish them from the subject of this study, phonetic vowel reduction.

One phonological type of reduction is reduction in vowel phoneme inventory in un­stressed positions: fewer vowel qualities contrast there than in stressed positions. Exem­plars range from Russian, which has one less vowel phoneme in unstressed positions, to most dialects of English, which are thought to have just one reduced vowel phoneme, /a/. Those Southern U.S. dialects discussed by Sledd (1966) are argued to have two reduction vowels, /a/ and /i/; /i/ may be considered the unstressed allophone of /i/. Bailey (1985:185) lists six unstressed vowels, [i, i, œ, ə, ɪ, ʊ], which may be considered as unstressed allophones of various stressed vowels. In all these analyses there is neutralization of contrasts in unstressed positions.

Another dimension of phonological vowel reduction is phonological mid-centralization in unstressed positions. English is again at one extreme, where /a/ is non-low and central, while Icelandic, with three peripheral “reduced” vowels /i,a,u/, is at the opposite extreme.

A third type of “reduction” is simply deletion of one out of a sequence of (phonologically identical) vowels, as discussed, for example, in Kenstowicz and Kisseberth (1979:128). This perhaps unrelated process will not be discussed further.

Finally, certain languages are said to lack vowel reduction (eg., Hindi and Spanish, cf. Dauer 1983), though the phonetic facts are relatively unknown (but see Delattre 1908). In summary, languages may or may not have either inventory reduction of vowel phonemes or phonological centralization of vowels in unstressed positions.
4.6.2 Phonetic Vowel Reduction

Phonetic vowel reduction refers to phonetic effects on vowels of reductions in other phonetic dimensions — that is, to the changes in phonetic vowel quality associated with decreased stress, sonority, duration, loudness, or articulatory effort. Vowel quality is a crucial phonetic variable because the identity of a vowel phoneme is perceived via the phonetic quality of its phonetic realization. The effects of reduction on acoustic vowel quality are largely unknown (though, once again, see Delattre 1968). Particular proposals exist for certain limited circumstances, but no general picture of reduction effects has been given.

One proposed effect of reduction is the assimilation of vowel quality to context, often called the undershoot effect. In this view, if adjacent segments are rounded, a reduced vowel would also become rounded (lowering formant frequencies, shifting back and up in (F1 vs. F2) vowel space); if fronted velar segments are adjacent, the vowel is front-velarized (shifting toward a high, front [i] quality). In a study of Stockholm Swedish short vowels, using nonsense syllables in fixed phrases repeated by a single speaker, Lindblom (1963a, 1963b) showed that vowels underwent contextual assimilation of vowel quality as duration reduced. He argued at the time that his findings failed to support the widely held theory that vowels centralize in quality when reduced in stress, since the effects could be entirely accounted for as a function of duration, which is itself partly function of stress.

Lindblom basically found that as a speaker produces a syllable faster and faster, the articulators simply cannot move fast enough to reach target positions for the consonant and the vowel. The result is that as vowel duration decreases, formant nuclei shift away from the “target” frequencies, towards the patterns characteristic of the phonetic context. This undershoot phenomenon is an example of “hard” coarticulation, which is governed by physical limitations of the vocal tract, such as the mass of the tongue, the distance between the vowel and consonant articulations, the amount of energy that can be expended by tongue muscles, etc.

Vowel undershoot of this kind is a phenomenon of general phonetics: when a syllable is so short that the tongue cannot physically move to a vowel target before a return gesture begins, then undershoot occurs. This universal physical constraint holds no matter what the language is. On the other hand, the whole range of linguistic fast-speech phenomena
may be thought of as planned ways to avoid reaching the physical limits of the speech articulators. It may be expected that such phenomena extend to vowels, so that planned simplifications of articulatory activities may be incorporated in faster or more relaxed speech. If this is the case, then vowel undershoot may be relatively insignificant among the vowel shifts that occur across various contexts. Thus, it is possible that there are linguistic forms of phonetic vowel reduction which do not follow simply from the brute physical sluggishness of the tongue. In such cases, the tongue may arrive at different inflection-points (phonetic nuclei) on its trajectory when stress or other environmental factors are varied, according to linguistic intent rather than physical constraint. It is quite likely that there is room for language-specificity in the vowel shifts associated with shortening and stress reduction, etc. Such phenomena belong to linguistic phonetics. They militate against extending Lindblom’s result to conversational speech and to other languages. But this is an empirical matter, which we will investigate in this thesis.

Phonetic vowel reduction was originally described as a process of centralization: “The average effect of consonants on vowels” is “a centralizing effect...all vowels are on the average shifted toward [ə] in contact with consonants.” (Joos 1948, quoted in Lindblom 1963b). Later it was described in laboratory studies as both centralization and contextual assimilation (e.g., Stevens & House, 1963). But Lindblom’s study found that centralization occurs “only insofar as the immediate context contains schwa [central] elements” (p1780) — that is, only through assimilation, or coarticulation. In short, reduction in duration had the effect of increased coarticulation. For this reason, future research should also consider the coarticulatory effects of adjacent segments, examining the extent to which increased vowel reduction is simply increased coarticulation.

Another, possible effect of reduction is phonetic shift along the path of ongoing sound change. For example, Labov, Yeager & Steiner (1972) found that upward-shifting /æ/ in polysyllabic words (i.e., in relatively short, unstressed syllables), was raised even higher than the most advanced fully-stressed /æ/ realizations. An early, small pilot study for this thesis (Mazzie & Veatch 1986) found a similar result: in the speech of a male adolescent Southampton (England) speaker, unstressed vowels were more distant from conservative forms than stressed vowels. One may speculate in this case that unstressed vowels have a quality advanced along the course of sound change. That is, a vowel’s distinctive quality
isn’t reduced if it lacks stress, rather it tends to occur farther along the trajectory of a vocalic sound change than its stressed co-allophones. The small amount of data examined in Mazzie & Veatch (1986) could be no more than suggestive.

Vowel quality could conceivably shift in any direction in vowel space under the influence of reduction, or none at all. Lindblom (1963b) points out that in Jassem’s (1959) study of Polish stress, “no relationship between stress and vowel quality could be demonstrated.” LYS and the Mazzie & Veatch pilot study for the current research suggest that other kinds of reduction besides simple consonantal assimilation or centralization might exist. Indeed, given the great variety in phonological reduction, one might expect variety in phonetic effects as well. The actual differences between dialects in the effects of stress discovered in this thesis are discussed in each of the dialect chapters, and summarized briefly in the concluding chapter, section 10.2.
Chapter 5

Methods

This chapter describes the kinds of data analysed in this thesis, how they are generated, and how they are to be interpreted. A number of important methods and fundamental methodological issues in phonological and phonetic research are discussed, including some new methods.

The phonological methods used in Chapter 3, Phonological Preliminaries, are briefly discussed. Then, the procedure by which each idiolect was phonetically analyzed is described, step by step. A method of impressionistically classifying vowels for their level of phrasal stress is given in detail. The clitic/non-clitic distinction is defined. This enables the exclusion of forms that may be phonologically reduced, in cases where this makes the pattern more clear. I discuss the criteria and rationale behind the choice of a particular time slice within the acoustic vowel as maximally representative of the phonetic quality of the vowel nucleus. Numbers of measured, excluded, deleted, and total tokens are given. A method of locating and accounting for outliers is described, along with some typical results of applying this method: most cases of apparent outliers are legitimate cases of extreme phonetic variation. The final section discusses statistical methods, including the bootstrap technique, which is used to create easily-interpreted displays of large amounts of data, as well as the statistical tests which are used to show that sets of F1-F2 measurements are significantly different.
5.1 Phonological Methods

In the chapter, Phonological Preliminaries, formal phonological analysis is applied to the problem of describing and explaining the vowel inventory of English. The approach taken attempts to explain the surface distribution of phonological features without reference to morphological factors. This approach may well be wrong. Most of phonology assumes that the regularities of the phonological surface are epiphenomenal, the result of the application of various processes to underlying morphological forms.

In the study of syntax in the 1970's, the role of surface structure become increasingly central, so that today, deep (syntactic) structure (in those theories that retain it), logical form and semantic interpretation, and phonological form are all related directly to surface structure. A similar tendency has not gone as far in phonology. Proposals have been made of "upside-down" phonology (Leben), where the more abstract forms are derived from the surface phonological structure, rather than the reverse. And recent work has emphasized well-formedness conditions on phonological representations, including conditions that hold true on the surface.

The present work carries this trend to a logical conclusion, and presents a grammar of some aspects of the surface phonological structures directly. It leaves open how other aspects of the system, such as the representation of consonants and the distribution of stress, are to be handled formally, as well the phonological aspects of morphological derivation, inflection, morpheme structure constraints, and so forth.

The approach taken here derives from two concerns. First, the crucial phonological question of this work is, What are the surface phonological representations which are the input to the phonetic interpretation system? Phonetic interpretation can be made sense of only when the phonological features and forms that are phonetically interpreted are first established. Thus it was necessary to describe surface phonological structure. Second, there are interesting phonological generalizations which are true on the phonological surface in English.

The basic forms of linguistic evidence used here are contrast and complementary distribution. Also used are intuitions about syllable-counting, and observations of various
historical and synchronic phonetic processes, as evidence for abstract phonological representations. Contrast is used to determine the presence of phonological distinctions (which need not be underlying in the lexical phonology, but are present in the post-lexical, or surface phonology). Complementary distribution is evidence for two conclusions at the same time: first, that the two items which are complementarily distributed may be the same thing at a deeper level, and second, that there is some process or alternation through which the surface distinction between the two is derived. For example, the phonetic vocalization of /l/ in some dialects is used to infer a phonological restructuring of the vowel system, which may be born out through predictable consequences for phonetic form and phonological contrast (page 50).

This thesis thus provides a distinctive approach to English post-lexical phonology, somewhat independent of lexical (or morpho-) phonology. It shows that phonology can derive interesting, significant results by observing patterns on the phonological surface.

5.2 Data

The primary data for this thesis is conversational speech taken from tape-recordings of sociolinguistic interviews of working-class native speakers of the four dialects studied.

The dialects were chosen to be very different from one another. They include exemplars of each of the “Three Dialects of English” (Labov 1991) plus a fourth dialect which is quite different from any of these and lies well outside this classification. The Three Dialects, so-called, are the Northern Cities, Southern, and Low-Back Merged dialects. The particular communities of the speakers studied here are Chicago, Anniston (Alabama), and Los Angeles (Chicano), while the maximally different dialect is Jamaican Creole.

A second reason for the choice of these particular dialects is the availability of the data. Four of my colleagues kindly gave me access to tape-recordings of some of the best interviews which they had made in the course of fieldwork in these communities. I wish to express my thanks to them here. The interviews analyzed are parts of larger projects on various linguistic aspects of the various dialects. The Los Angeles Chicano data is part of Otto Santa Ana’s (1991) dissertation on various phonological issues in that dialect, including cluster simplification, as well as vowel reduction. The Chicago interviews by
Sharon Ash are part of the project on Cross-Dialectal Comprehension, another interview by Benjamin Wald is part of the data on which Labov, Yaeger and Steiner (1972) was based (though acoustic analysis of this speaker, Jim C., was not done there). The Anniston, Alabama, material comes from interviews done by Crawford Feagin in 1971 as part of the fieldwork behind her (1979) book, Variation and Change in Alabama English, and which she has drawn on in other work since then (1986, 1987, 1990, and to appear). Finally, the Jamaican Creole data is from sociolinguistic interviews done by Peter Patrick, who is completing his dissertation, a sociolinguistic study of urban Jamaican Creole in a neighborhood of Kingston.

The relation of the interviewer to the interviewee is partly controlled in this data. The interviews are all same-sex conversations (excepting Feagin's interview with James H.). The interviewers are in all cases trained sociolinguistic fieldworkers who grew up in a speech community near to the one studied. Otto Santa Ana is himself a Chicano American from Arizona, so while he is not a native member of the Los Angeles Chicano speech community, his personal background makes him much better able to get natural behavior from LACE speakers than an Anglo such as myself might be. Crawford Feagin was raised in Anniston, Alabama. Sherry Ash is herself from the Chicago area. And Peter Patrick spent his first 15 years in Jamaica and is marginally a native speaker of JC. In every case, the interviewer was a somewhat marginal member of the speech community studied, neither an insider nor an outsider.

Considering that non-working-class academics are generally outsiders to the working-class communities near where they grow up, and even more so far from home, this data is about as natural as can be gotten in this type of interview. Certainly interviewers without detailed knowledge of the community studied would have stimulated a very different set of behaviors, that are likely to be much less characteristic of natural conversation between native speakers of the particular dialect.

The interview situation is also controlled: each is a sociolinguistic interview. (For details, see Labov 1984:32-42.) The format of the sociolinguistic interview is designed to get the maximum quantity of unmonitored speech, preferably in the form of narratives of personal experience, from a single speaker, in a fairly short period of time. The sociolinguistic interview is a one-on-one conversation, with the interviewer leading a rather open-ended
conversation along paths that appear to be interesting to the speaker. A good interviewer knows what topics are likely to be interesting to their subjects. The desired (and in these cases, attained) result is that the interviewee talks much more than the interviewer.

The equipment used to record the interviews consisted of either portable reel-to-reel 1/4" tape-recorders (for Chicago and Alabama speakers), or an excellent quality cassette tape recorder (for the Jamaican Creole and Los Angeles Chicano speakers). Sound quality varies from excellent to good, though sometimes overlapping speech and intermittent background noises (particularly for Jim from Chicago, and Roasta from Jamaica) required particular words or phrases to be excluded.

The data that is derived from these tape-recorded interviews includes all the measurable vowels that occur in lengthy parts of the interview. The data thus includes both stressed and unstressed tokens and a range of consonantal contexts, which are two of the variables that will be focused on. It should be noted that studies of less-monitored speech are more likely to show increased frequency and degree of application of phonetic processes such as coarticulation and reduction. An important disadvantage of this method of data analysis is that the sample is skewed by the token frequency of the various features of interest. Even the law of large numbers, by which bias is minimized as the sample size increases, may not help, since the token frequencies may be systematically skewed by certain high frequency words. This disadvantage can be minimized by careful attention to the skewing factors. On the other hand, while the data may be skewed in some ways, it is real speech data, which relatively closely reflects the actual vernacular speech patterns which speakers of these dialects use when communicating with each other.

5.3 Analysis Procedures

5.3.1 Background

Below I discuss how data may be transformed from tape-recordings into summary files containing phonological and linguistic categories and phonetic measurements which are the empirical basis of the present studies of phonetic implementation. Some readers may find the description of this transformation to be overly detailed, but I include it in the belief that some will find it helpful.
Some background for this discussion is necessary. An important stimulus for this work is technology. The study of the sounds and sound systems of language has benefited repeatedly from advances in speech technology. Convenient field recording and reproduction of sounds was impossible until the tape-recorder; rapid broad-band spectral analysis was impossible until the Sound Spectrograph; and consistent and precise automatic estimation of vowel formant frequencies was unattained until the development of Linear Predictive Coding.

Software advances have enabled easy, interactive data collection, analysis, and display. With mouse-driven interaction with graphic displays of data, tasks like the extraction of formant frequencies from vowel nuclei can be done at a rate of many vowels per minute, compared with the minutes required for each vowel token using earlier technologies. The continuing efforts of a community of electrical engineers has made possible relatively accurate, automatic extraction of formant and pitch contours.

Hardware advances have come on at least two fronts: processing speed, and mass storage size. Computer processing speeds have reached the point where spectrograms can be calculated in near real-time on relatively affordable workstations. Thus if the scientist wishes to examine waveforms or spectrograms or formant tracks for any particular segment of speech, this is no longer a time-consuming project requiring thought and dedication to the process of data analysis itself, but just a few seconds of work. Research now becomes more a matter of immediately testing one’s ideas rather than painstakingly manipulating low-level details.

The computer’s capability to store and access large quantities of data has grown qualitatively, changing the kinds of work that may be done. Hours of speech can be stored on-line and accessed randomly. This allows phonetic studies to examine whole conversations, which previously were too large to be stored on a single computer.

Consider an example central to the phonological discussion above: if one phonologist could effortlessly listen to and examine spectrograms of numerous instances of /ayr/ in natural speech for different speakers and different dialects, then the objective reality of the intuition-based claim of another phonologist (in this case, me) that this sequence is bisyllabic could be assessed with ease and certainty. Or if I wonder about the phonemic content of the word “get” for Jim from Chicago, I can issue a short command to a computer, to
locate all twelve tokens in the transcription and in moments listen to each of the utterances twice: the [i] quality common to them all suggests that it is /giʃ/, not /gɛʃ/. Checking the hypothesis with /i/ and /ɛ/ in other environments, with other speakers, etc., takes only moments more. Examples could be multiplied further. Phonological and phonetic research practices may be revolutionized by these technological advances.

5.3.2 A Twelve-Step Program for Acoustic Analysis

A 12-step AA program is described here, which I used to convert a tape-recording into a summary file of phonological and acoustic information characterizing all the vowels in a discourse, suitable for input into a statistical software package for data display and analysis, which is used to generate the charts and statistics that form the essence of this thesis. I hope that this program will be useful to others, but also that they will not simply accept it unquestioningly.

The steps in this procedure require very different amounts of time. Some take only seconds (steps 8, 9, 12), while the others are quite slow. The central step in the procedure, that of locating nuclei and correcting erroneous formant-tracks (step 10), is the most labor-intensive part, despite the improvements in efficiency of an order of magnitude or so, introduced here.

(1) Suitable sections of the tape-recording were located, by finding the most animated, un-self-conscious speech, preferably narratives, for a total of 6 to 10 minutes of suitable speech (but 25 minutes were taken for the first speaker, Rosie S. from Chicago). This results in roughly 2000 vowels’ worth of speech for each speaker (exact numbers in Table 5.2 below). This unit of speech closely matches the definition of the “idioclect”, by Bloch (1948): The speech of one speaker speaking on a single topic for a short period of time.

(2) The desired section was digitized at a sample rate of 12,000 Hz, with a sample resolution of 15 bits.

(3) The resulting large waveform file was broken into smaller chunks separated by periods of silence. This unit of speech was defined by Bloch (1948) as the linguistic “utterance”. Break points are typically located in the middle of silent periods, but occasionally when the period between silences contains an unwieldy amount of speech, a break-point was be
inserted between intonational phrases even if they are not separated by definite periods of silence. For example, if there appears to be phrase-final lengthening and drop-off in pitch associated with the end of a sentence or other large phrase, then I sometimes insert a break-point between that and the beginning of the immediately following sounds. Similarly at the end of a false-start a break-point may be inserted. Each such delimited segment will be called an “utterance”.

(4) Each utterance was formant-tracked.\(^1\) Formant-tracking subjected the signal to a number of operations. The entire waveform was downsampled to 10,000 samples/second and high-pass filtered to remove DC and low-frequency rumble. LPC poles were computed 100 times per second, using 12th order autocorrelation LPC, with a pre-emphasis factor of 0.7, and an analysis window of 49ms, weighted by a cosine\(^4\) weighting function. While the window is relatively long, the weighting function reduces its effective width considerable, relative to a rectangular window. One Chicago speaker, Rita, was analysed with order 10 autocorrelation LPC, pre-emphasis factor of 0.85, to optimize the LPC analysis and formant-tracking parameters to her speech. The remainder were satisfactorily tracked using the above default values. Not all the LPC spectral peaks represent formants: some of them are used to model the overall shape or tilt of the spectrum, while others are located on spectral peaks which may be identified as spurious or temporally discontinuous or which have a wide bandwidth). A post-processing technique\(^2\) is therefore used to eliminate the spurious LPC peaks, and label a fixed number of formants (the default number is 4; 3 were tracked for Rita).

(5) The entire discourse was transcribed orthographically (except for Jamaican, which was transcribed directly in phonemic form).

(6) A dictionary of orthographic words and their phonological forms was created. Many or all instances of each word were listened to, in determining the phonological form(s) of the word. The orthographic transcription was converted into a phonological transcription according to the dictionary. This is done by a conceptually simple program written for

\(^1\)The computer software used in this project for digitizing, for formant-tracking, for display of waveforms, spectrograms, and formant-tracks, and for segmentation, are parts of the waves+ package developed by David Talkin at AT&T Bell Laboratories, and available commercially from Entropic Speech, Inc., Washington, D.C.

\(^2\)This post-processing technique is described in Talkin (1987). Cf. also Secrest and Doddington (1983), and Dupree (1984).
the purpose, which replaces each word's orthographic form by its phonological form and inserts word boundaries. Morphophonological alternations, phonological external sandhi effects (such as /ðə/ vs. /ðiː/, etc.), are accounted for by providing different phonological forms for certain words.

(7) All the vowels were impressionistically coded for phrasal stress, as described in Section 5.4, below.

(8) All the vowels were extracted from the transcript, along with their stress levels and their adjacent consonantal and vocalic contexts. This was done automatically by a program written for the purpose.

(9) Each vowel was classified as occurring in a clitic or non-clitic word, according to the criteria given in Section 5.5 below, in order to be able to exclude tokens whose surface phonological forms may be indeterminate due to phonological processes of reduction.

(10) The temporal locations of the onset, nucleus, and offset of each measurable vowel was marked. This step is the most painstaking and time-consuming one, taking about an hour to do 150 vowels, and a day to do 600 (at a sane pace, though greater rates are possible in spurts). Each utterance's waveform is displayed on a computer graphics terminal; a broad-band spectrogram on an expanded time- and frequency- scale is computed using special signal-processing hardware which does the task in near real-time (that is, calculates a spectrogram of one second of speech in about one second). The raw LPC peaks as well as the automatically-classified formants are overlaid in color on top of the gray-scale spectrogram. The user is prompted to specify a temporal location for each of the segmentation marks, by clicking a button on a mouse-controlled pointer on the screen. At this point also, the formant tracks are examined to see if they correspond to the auditory quality of the sound and to check that they follow real resonances shown in the spectrogram. Both the raw LPC peaks and the automatically-classified formant tracks are overlaid directly on top of the spectrogram. This allows for immediate visual checking of the correspondence of the automatically-measured formants to true resonances that are evident on the spectrogram. Sometimes mistracking occurs; a formant measurement may occasionally appear at a point where there is no energy in the spectrogram, (for example, between two resonances). When mistracking occurs, there is almost always some LPC
peak at the correct frequency location, which has been incorrectly ignored by the post-
processing algorithm. For this situation, the software provides a facility for modifying the
labelling of LPC peaks. One selects a formant with a mouse-controlled pointer, and draws
the pointer over the correct LPC peaks, and those peaks are relabelled as formants. Thus,
formants can be re-associated with LPC poles that lie on visible spectral peaks in the
spectrogram. This procedure is used to correct all the erroneous formant tracks, during
a detailed, vowel-by-vowel examination. The onset and offset of the acoustic vowel are
located at acoustic discontinuities where they occur (with a precision of around 3 mil-
lieconds) (see the definition of “acoustic vowel” in Chapter 4) and at points of maximum
spectral change where they do not. The nucleus of the vowel is located according to the
methods in Section 5.6 below, which discusses nucleus-picking. If there is no acoustic
vowel corresponding to a particular vowel phoneme, it is not marked. If there is over-
lapping speech or other noise, distorting the formant measurements at any point in the
syllable, that vowel token is thrown out, deleted from the corpus. The total numbers of
unmarked, deleted, and measured tokens are given in Section 5.7 below.

(11) Extract the formant frequencies from the corrected formant-track files correspond-
ing to the time-locations of the marked nuclei.

(12) Collect all these pieces of information together into a grand summary file, which
contains the following 16 pieces of information in columns.

• a unique token ID number,
• the vowel phoneme,
• the preceding and following vocalic and consonantal contexts (4 character strings),
• the impressionistically coded phrasal stress level,
• the clitic vs. non-clitic classification,
• the phonological form of the word in which the vowel occurs,
• the file containing the waveform for that utterance,
• the time of the onset, nucleus, and offset of the acoustic vowel,
• F1, F2, and F3 at the marked nucleus time.

There is some large number of lines per file (exact numbers given in Table 5.2 below), one line for each token measured for the speaker in question. This summary file becomes the input to a data display, analysis, and programming package called S. The various charts, statistics, and other results given in the body of the thesis are generated by operations on the data within that statistics package. For example, an F1-F2 plot of all measurements for a speaker is created by the command, plot(-spkr$f2,-spkr$f1). (Remember that plotting -F2 vs. -F1 gives the traditional orientation of vowel charts, with [i] at the upper left, [u] at the upper right, and [a] at the bottom in the middle.)

5.4 Impressionistic Coding for Phrasal Stress

For the purposes of this thesis, stress is taken to be an intuitive classification of syllables as relatively more and less prominent within a spoken utterance: impressionistic classification of phrasal stress. This is not a pattern of stress values predicted from the structure of the message, but a subjective coding of impressions of prominence. For this reason, results may be somewhat different from, e.g., Crystal and House (1990) who used lexical stress classifications. Phrasal stress has a strong effect on vowel quality; in order to see what these effects are, it was necessary to code each vowel for its degree of stress.

While intuitions about lexical stress and stress patterns in complex nominals and clitic groups are quite clear, it is well known (Lieberman 1963) that the intersubjective reliability of phrasal stress coding leaves something to be desired. My own reliability tests described below found >10% disagreement using three levels of stress; four levels may increase intercoder inconsistency to 35% (Mazzie & Veach, 1986). This kind of observer "noise" is probably characteristic of all impressionistic phonetic judgments per se. (This is not true, of course, for judgments of phonological or lexical identity, which are extremely reliable.) To some extent this reflects issues of definition and of training. There may well be deeper problems, a matter that we cannot resolve here.

Here, as clearly as I can define it, is the procedure I used to code phrasal stress, expressed as a set of instructions to a coder.

Before coding for stress, listen a few times to a few minutes of speech, and try to listen for the rhythm of her/his speech. You will be more consistent if you have an intuitive feeling for what that speaker’s rhythm sounds like.

Listen to entire utterances (breath groups) rather than single syllables, so that you can hear the overall rhythmic contour. Try to keep the sound of the entire utterance, or many-syllable parts of it, in your mind at one time. Then locate the most prominent syllable(s) which seem to carry the primary beats of the utterance. Don’t concentrate on the sound of an individual syllable, but listen to the whole utterance and pick out the prominent parts or the beats. Pay particular attention to prominences marked by pitch inflection. If beats do not stand out in your mind, then try repeating the utterance to yourself in a monotone, without melody, while overemphasizing the rhythm, as if counting meter in poetry. Put the beats of the rhythm where they sound right to you (that is, where you would put them), even if you’re not sure the utterance itself had that pattern of beats. Then listen back to the utterance for confirmation or denial of your intuitive analysis. As you listen to the utterance, if what you hear conflicts with the rhythmic pattern in your mind, then change the rhythm in your mind, to one that is in accord with the actual utterance. Repeat this until you have something that fits. Try to find the best interpretation, even though there may sometimes be more than one interpretation that fits.

When you have a rhythmic pattern for the entire utterance in your mind at once, all the beats in that rhythmic contour are primary beats, coded with stress level 1. Then if there are other syllables that seem strong in addition to the primary beats, for example, when paying attention just to shorter phrases, code them as secondary stresses (level 2). The remaining syllables are “unstressed”.

This method is most difficult to apply when pause-groups are very short, from one to four syllables. If the entire utterance is pronounced weakly, one may code the strongest beats as secondary stresses. If all the syllables seem prominent, one may try to separate out the pragmatically meaningful intonation contour from the stresses, and consider that some of the prominence of the syllables may derive from the intonation contour’s requirements.
For example, a high pitch associated with the beginning of a particular intonation contour will be realized on the first syllable in the utterance even if that syllable is unstressed.

There are special concerns to be remembered in coding utterance-final syllables. Pre-pausal lengthening applies to unstressed syllables, as well as stressed ones. Thus in long and short utterances, the last syllable may be lengthened despite being unstressed. Final lengthening does not seem to apply to the last syllable in a false start or a hesitation – this may be the primary phonetic “cue” to the “editing signal” of Labov (1972:203) and Hindle (1983) associated with false starts. One may allow for some lengthening in final syllables, taking into account the effects of the editing-signal: these are not necessarily cues that the final syllable is stressed.

Another procedure that one might have followed in coding the stress contour is to decide the relative prominence of each syllable by making a forced-choice among stress levels, moving one syllable at a time from left to right in the utterance. This method, however, is subjectively very difficult, and it may give unstable results. I found it much easier and more robust to listen for the globally most prominent syllable(s) of the entire utterance.

The analysis is thus, loosely speaking, top-down rather than left-to-right. A final consequence of this method is that primary- and zero-stressed syllables are more common than secondary-stressed syllables, since secondary-stressed syllables are coded in a sort of afterthought.

Two consistency tests were conducted. One examined how consistent I was in coding the same material on two separate occasions; this is discussed in the chapter on Jamaican Creole, page 182. A second, small consistency test compared two coders. This test was conducted on the second Chicago speaker, Jim C. After ten minutes’ training and discussion with a recent graduate of the 1st-year phonetics class, 128 syllables were classified for three degrees of stress. The results are displayed in a confusion matrix in Table 5.1.

There are six 2-level inconsistencies; the number of 1-level inconsistencies is 9, and the

---

4Interestingly this is not a positive cue, but rather the non-application of a regular process, which is in a sense a negative cue. Another cue to the presence of this “editing signal” is phonotactically impermissible syllable endings, e.g., *str-. There are also explicit editing-signal morphemes, such as uh, um, etc. I speculate that another editing-signal cue is an apparent stress on closed class items which shouldn’t be stressed, where the appearance of stress comes from a rapid change in the pitch contour at the boundary of the hesitation.

5William T. Reynolds.
Table 5.1: Confusion matrix of phrasal stress classifications by two coders.

<table>
<thead>
<tr>
<th></th>
<th>Mine Primary</th>
<th>Mine Secondary</th>
<th>Mine Unstressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>28</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>His</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Unstressed</td>
<td>2</td>
<td>4</td>
<td>82</td>
</tr>
</tbody>
</table>

consistent classifications number 113/128 for an average of 88.3% cross-coder consistency. This is comparable to the 85% level of intercoder consistency found in Mazzie & Ventch (1986). They found only 65% consistency when coding for four levels of stress. If the secondary-stress level in this data is collapsed together with either primary stress or unstressed, leaving only two levels of stress, then the percentage of inconsistency is reduced from 11.7% to either 8.5%(11 cases, secondary classed with primary stress) or 7.8%(10 cases, secondary stress classed with unstressed). Neither is significantly better than the other, and in the stress data used below, I have collapsed secondary stress together with primary stress, as against unstressed, to form a binary, stressed-vs.-unstressed classification of syllables. Using these results to make an estimate, the degree of inconsistency across coders, making a forced-choice, binary classification, is in the range of 5% to 10%. The self-consistency test, discussed in the Jamaica chapter, page 182, was conducted on 231 syllables from one of the Jamaican speakers, with myself doing the classification on two occasions, nine months apart. The results are indistinguishable from these: 87% consistency using three levels, 92% consistency using two levels.

There are a number of sources of these inconsistencies. Some of them are simple errors, due to lack of attention, or carelessness, while some inconsistencies come from other, more interesting sources. For example, it seems to me that I can often change the stress pattern I hear on a given sentence just by actively thinking of it in a different way. Both subjectively-perceived patterns seem consistent with the utterance itself as I listen to it. This is an unfortunate source of ambiguity.

Finally, the usual cues for stress (pitch inflection, duration, and loudness) may not

---

*Dodginton estimates usual human error rate in making categorical classifications of clear cases at around 4%. However when highly motivated, the level improves to under 0.5%. In the current case, the motivation of the coder (me) is undoubtedly greater than that of a poorly paid, unmotivated experimental subject, but not as great as the financially highly motivated subjects in Dodginton's study.
always be correlated in a given utterance. Extra-long syllables (even in non-final position) may be melodically monotonous and lack a pitch inflection. If in a given sequence of two syllables, the first syllable has a sharp pitch inflection but is quite short, and the second is long and drawn out but has no definite rise or fall of pitch, the different cues lead to different classifications. If the coders are listening for different cues, they will be inconsistent to the degree that the cues to "stress" are inconsistent. While many cases are clear and the cues are strong and mutually consistent, many other cases are unclear, to such an extent that is is worth seriously considering whether stress is nothing more than the subjective interpretation of linguists. However, as we see in later chapters, this subjective interpretation correlates well with significant patterns of changes in vowel quality, as reflected in F1, F2 measurements.

5.5 Clitics Versus Other Words

All lexical items in the corpora were coded as clitics or non-clitics according to the following definition, which was adopted for the purpose of distinguishing vowels whose phonological category is variable or uncertain from those which are more clear. This classification becomes useful later when the phonological variability of clitic words becomes an impediment to the analysis; clitic words items can then be excluded from the analysis.

Clitics are monosyllabic, closed-class words which are capable of having completely reduced vowels. For present purposes, the class of clitics was defined as the set of typically unstressed and monosyllabic grammatical words, plus added contractions. These are pronouns (optionally plus contracted modals: 'd 's 're 're); non-numeric determiners; and unstressed prepositions (stressed prepositions are not included, such as along, between, about, before, etc.); and forms of the auxiliary verbs be, do, have, will, and can (optionally +n't). It should be noted that, contrary to general opinion, negated auxiliaries (such as isn't, wouldn't, can't, etc.) reduce to some extent in some cases, since they are not always the focus of the sentence they occur in, and therefore they can also be questionable tokens. Thus I included them in this category of reducing forms.
5.6 Locating Acoustic Nuclei

The methods used here to choose a time slice of the acoustic vowel as representative of the vowel's nucleus are similar to those of Labov, Yaeger, and Steiner (1972:29) (LYS). Typically the maximum of the F1 contour within the acoustic vowel is chosen. This time-point reflects the point of maximum opening of the mouth (cf. Chapter 2, page 21), and also the target or point of inflection of the opening gesture of the vocal tract which articulatorily implements a phonological vowel. As the vocal tract moves from consonant to (simple) vowel to consonant, it opens and closes. If there is an articulatory or acoustic target associated with a vowel, the point of maximum opening of the mouth is commonly the point at which that target is most closely realized. There may be no actual target or "magic moment" within a syllable that constitutes the essential realization of the vowel nucleus; instead the entire trajectory of the gesture associated with a sound may be an essential part of the sound. Even here, the point of maximum opening is of great importance, since these trajectories may be most simply characterized by describing the acoustic or articulatory configuration at this point, along with the time-course of the transition to this point from preceding segments and from this point to following segments. This is the technique of "temporal decomposition" (Atal 1983), a mathematical method in which the actual smooth sequence of overlapping gestures is decomposed into a sequence of separate targets and a set of time functions which describe how they overlap and shade off gradually from one into the next. This method may be useful to gesture theorists (cf. Browman & Goldstein 1986) in the mathematical characterization of the entire trajectories of speech gestures.

It is easy to delimit and listen to very short pieces of the speech waveform. I often listened to a short segments (as little as 30ms) which included the time-slice marked as the acoustic nucleus of the vowel in question, as well as larger segments containing the entire vowel and adjacent consonants. It was my frequent impression that the short nuclear segment sounded most like the overall impressionistic quality of the entire vowel, as compared with earlier or later short segments. Even segments adjacent to the acoustic nucleus in lengthened vowels did not generally sound as much like the entire vowel as the segment centered on the F1 maximum. Thus the F1 maximum seems to be an auditorily important
point in the course of the acoustic vowel.

If F1 has no maximum, then the point of inflection of F2 (maximum or minimum) is chosen, considering that the gesture that created the sound reached an inflection point at that time-point, and began to move in the direction of other targets. If neither has a point of inflection, then the mid-point of the acoustic vowel is chosen. These considerations must be altered in the case of downward-gliding vowels, such as ingliding, raised /æ/ [æ\textsuperscript{L-2}], /ir, i:l/ [iə, əl], Jamaican /uo/, and the like which show formant trajectories in which F1 rises and F2 falls between the part of the acoustic vowel that corresponds to the vowel nucleus and the part that corresponds to the following glide. In such cases, the middle of the section that corresponds to the phonological nucleus of the vowel is chosen as the acoustic nucleus; commonly this is a steady-state preceding the offglide.

Labov, Yaeger, and Steiner (1972) considered only stressed monosyllables, while this study includes all vowels occurring in a lengthy discourse, from the shortest to the longest, in order to include all vowels, stress effects, and consonant contexts. Thus the restriction in LYS that the measurement be at least 50ms after the beginning of the syllable was abandoned here. If the acoustic vowel was 50ms or less in duration, the midpoint of the vowel was chosen.

5.7 Exclusions, Deletions, Total Measurements.

Within the segment of speech chosen, ranging from a few minutes to nearly half an hour, an attempt was made to examine every single occurring vowel. A small percentage of the vowels are devoiced or otherwise deleted in speech, while another, usually small subset was excluded due to background noises or overlapping speech. The overall numbers of tokens in these various categories are shown in Table 5.2. For each dialect and speaker, it lists the total number of vowels examined, the number of vowels which were found to correspond to no acoustic vowel (that is, those which were phonetically deleted, devoiced, etc.), the number of vowels which were excluded due to overlapping speech, or background noises, and the total number of vowels measured.

The percentage of phonological vowels which were found to be associated with no measurable acoustic vowel segment is calculated as the number of phonetically deleted
Table 5.2: Numbers of tokens, deletions, exclusions, measurements.

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Speaker</th>
<th>Total</th>
<th>= n(V→∅) + n(excl.) + n(meas.)</th>
<th>% (V→∅)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC Juba B.</td>
<td>2891</td>
<td>82</td>
<td>129</td>
<td>2680</td>
</tr>
<tr>
<td>JC Roasta M.</td>
<td>1872</td>
<td>75</td>
<td>124</td>
<td>1673</td>
</tr>
<tr>
<td>CWE Rita S.</td>
<td>4821</td>
<td>211</td>
<td>140</td>
<td>4470</td>
</tr>
<tr>
<td>CWE Jim C.</td>
<td>2856</td>
<td>121</td>
<td>406</td>
<td>2329</td>
</tr>
<tr>
<td>CWE Judy H.</td>
<td>1775</td>
<td>105</td>
<td>55</td>
<td>1615</td>
</tr>
<tr>
<td>AE James H.</td>
<td>1818</td>
<td>137</td>
<td>43</td>
<td>1638</td>
</tr>
<tr>
<td>LACE Vince M.</td>
<td>2190</td>
<td>104</td>
<td>196</td>
<td>1890</td>
</tr>
</tbody>
</table>

\[ \%(V\rightarrow\emptyset) = \frac{100 \times n(V\rightarrow\emptyset)}{n(V\rightarrow\emptyset) + n(\text{meas.})}. \]

segments out of the total number of measurable segments.

The deleted/devoiced tokens might be used as data for another study, examining the conditioning of vowel deletion and/or devoicing. My impression, as might have been expected, was that vowels deleted most often between voiceless obstruents. A more detailed characterization of these exclusions will have to await further analysis beyond the scope of this thesis.

5.8 Outlier Analysis

One problem in large-scale analysis of this sort is that erroneous measurements may slip by, due to insufficient attention. Much of the phonetic variation documented in this thesis is quite extreme, much more extreme than what is found in monitored “laboratory” speech. A skeptic could wish to attribute some of the apparently unusual measurements to mistakes of measurement. This section attempts to assuage the doubts of such skeptics, first by showing how apparent errors (outliers) were found, and second by showing how most of these apparent errors are in fact correctly measured phonetic forms whose characteristics result from natural phonetic and linguistic influences.

The error-detection procedure followed is to plot an F1-F2 chart on a computer screen, displaying the formant-measurements of a single vowel class and speaker at a time, and then to identify any gross outliers. A button-press on a (mouse) graphic pointing device next to an outlying data point on the display reveals the identity of that outlier. For each outlier, the waveform, spectrogram, formant-tracks, and segmentation times are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
redisplayed, listened to, and checked for mistakes.

In most cases, the outliers turn out to be correct measurements – a function of the considerable care used in step 10 of the analysis procedure above. Thus of the 13 most distant outliers identified in the 2329 measured tokens for the 2nd Chicago speaker, Jim C., 2 were errors due to a phonological misclassification of /ə/ as /æ/ in the words can and and, while 11 were correct measurements. None were actual errors of measurement. Among 20 and 17 outliers examined for the Jamaican speakers, 2 tokens each were found to be actual errors, while the others were due to extreme assimilation to adjacent sounds or due to extreme stress. When in the great majority of cases the most distant outliers are actually correct measurements, then the less-outlying tokens are even less likely to contain gross errors. Gross errors appear as outliers, and the remaining tokens are relatively consistent with each other and less outlying. The most extreme outliers are only rarely errors. Variation is in fact extreme, as is shown, for example, in Figure 7.4, page 202. Thus although some small residue of errors undoubtedly remains, this data is fairly clean.

How can outliers be correct? The following list of outliers in the speech of Jim (from Chicago) discusses why these 11 tokens were phonetically unusual, and gives the reader something of a flavor of the kind of striking but regular acoustic variation that occurs in natural vernacular speech.

- 120 /ow/ is very low. "No!", /now/, [na°]: This carries emphatic, contrastive intonation. It's not "nah"; this is clearly the word "No", because of the offglide [°].
- 201 /ay/ is well to the front of the other /ay/ tokens. “at night”, /æt#nayt/: The preceding front /æ/ seems to pull this token to the front.
- 794 /æ/ is extremely low and back. “blank” /blæŋk/, primary stressed. F2 glides forward from the preceding /l/ into the following front /ŋk/; the nucleus is chosen, as usual, at the F1 maximum. Following velars have a strong retarding effect on the raising of /æ/, in Chicago just as in New York City and Philadelphia. This effect appears magnified in this doubly lengthening environment: it is stressed, and also precedes a voiced tautosyllabic consonant.
- 1620, 1817 /uw/ very front. Both occur in “I threw the” (paper, bottle) /θruw#ðə/:
/ð/ is deleted. /0r/ is realized as a palatal affricate, and the entire realization of /uw/ is thereby strongly fronted.

• 1891 /i/ is high and back. "I knew if I lasted", /nuw#ɪf/: /i/ in the clitic form "if" (cf., "If you try it" = "Few try it") is reduced to mere length on the preceding /uw/. Its phonetic timing slot is not apparently lost, but its phonetic quality is entirely due to the adjacent stressed /uw/. This case could be analysed as compensatory lengthening of /uw/, or as feed-forward coarticulation.

• 1919 /uw/ is very low and front. "I knew I didn't", /nuw# aɪ/. The preceding coronal has the typical American effect of fronting the nucleus of /uw/, realized as a quite high F2 onset, gradually falling throughout the vowel for about 70 ms in this case. The lowering which also occurs may be attributed to the influence of the following low nucleus of /aɪ/.

• 2096 /a/ is the lowest token of all; it's off the chart. "... threw bricks at me and Holly" /hɔli/. The /a/ is emphatically stressed, slightly creaky, but the formants are quite clear, and the preceding fully-realized /h/ (without oral constriction) ensures that this low vowel cannot be attributed to coarticulatory influences. The conversational context is a list of events in which the speaker was attacked; once he was with his friend, Holly, who is 6'2", 200 pounds, and a person to be feared in a fight. Thus the affective stress on "Holly" symbolizes the irrational nature of the attack. The stress is realized by an extension of the mouth-opening gesture associated with /a/ (thereby raising F1, cf. Chapter 2, Acoustics) beyond what is found for any other token of that or any other vowel.

• 2330 /ɔ:/ is very front: "He thought he was.." /hɔɪ#ɔt#hiy/ /t#h/ is deleted. This realization of /ɔ:/ sounds lower mid, central, [v]. Here the surrounding high-front /iy/’s seem to have fronted this nucleus, though they didn’t raise it.

• 2466 /ɔ:/ is far lower and fronter than other tokens of /ɔ/: “All right” /ɔl+rayt/ /ɔ/ is the most stressed syllable in this utterance. This cannot be attributed to coarticulation with the following /ay/, since this /ɔ/ is even lower (but backer) than the following /ay/ nucleus. Instead this may be an effect of sound change in
progress, where /a:/ lowers and fronts (cf. the Chicago Loop Chain Shift, discussed in the Chicago chapter), apparently more so when stressed.

- 2764 /I/ is very low. "I got him down" /gat#IM/: /I/ in this clitic form has reduced qualitatively to a glide between /a/ and /m/, but has its own timing slot (measured at 90ms, which is quite long).

In this data, the formant-tracking was good to begin with, and the formant tracks were checked and corrected by hand for every vowel token. These examples show that extreme outliers on F1-F2 charts derived from these measurements are most often to be attributed not to errors of measurement, but to particular phonetic and linguistic circumstances that produce the variation in an understandable, even predictable, way. (Thus for example, not just one but both tokens of “threw” showed extreme fronting of /uw/.)

Freud said that to understand the normal, it is helpful to study the abnormal. The stressed tokens above which lay well outside the “normal” distribution of the respective vowel phonemes may be a window to the understanding of normal speech. In particular, it is clear that heavy stress can exaggerate articulatory gestures, giving rise to these outlying measurements. The hyper-open /a/ in “Holly” remains /a/, but its qualities are not those of other /a/’s: the speaker’s mouth was considerably more open. If the phonetic realization of /a/ for this speaker is a particular acoustic target in formant-space, this token is not a proper /a/, because it lies well outside the normal range of formant-frequencies. But it is a proper realization of /a/, in this particular linguistic context. It may resolve this paradox to say that vowels are realized by a particular articulatory gesture, which may be overlaid onto other gestures, or magnified as a result of stress or lengthening. Then the form of the gesture (e.g., mouth-opening) may remain while parametric modifications to the gesture (duration, magnitude, relative timing with respect to other gestures) may be made through the influence of context, stress, etc. An acoustic analog of this articulatory “gesture theory” (developed in various papers by Browman & Goldstein, e.g., 1986) may also be possible.

In summary, acoustically unusual outliers are frequently natural, sometimes even predictable realizations of sound-classes in particular environments, which may be used for getting deeper insight into the nature of speech sounds. Errors of instrumental analysis,
while undoubtedly present, cannot be used as explanations for the patterns of extreme variation found.

5.9 Statistical Analysis

The central statistical problem of this thesis is that of demonstrating whether the distributions of two sound classes in F1-F2 space are the same or different. Different distributions in this acoustic space are evidence for some linguistic difference, whether phonetic or phonological, as argued earlier. Therefore the problem of characterizing a difference as significant or insignificant is crucial.

If two categories of data are normally distributed, then well-understood analytical methods are available to test the hypothesis that the two categories are different. Such methods exist even for multivariate data, and even for log-normal distributions. However, formant-frequency data is frequently not normally distributed. For example, the distribution of /e/ in the Chicago chapter (page 202 is quite non-normal, as are any number of charts of raw formant distributions.

A test exists for testing the question, Do two distributions have different means? which does not depend on the normality assumption: the Wilcoxon test. But tests of this nature for multi-dimensional data are less well-known (though see Maekawa 1989 for a two-dimensional t-test, not used here). I use two methods in dealing with this problem. The first is used as a technique of visually displaying differences in a way that is easily interpreted. The second is used for numerically estimating the statistical significance of the difference between two sets of measurements.

5.9.1 The Bootstrap

Recent advances in statistics have developed extremely general, yet simple methods for estimating confidence intervals and other information about the distributions of sample statistics, using a resampling technique known as the bootstrap.7 The bootstrap method may be used in an attempt to estimate any aspect of the distribution of any statistic (Efron

7Equally well-known is a resampling technique known as the jackknife, so-called because it’s good for many purposes. The bootstrap is simpler and even more useful (though it will not quite allow you to pull yourself up by your bootstraps).
The technique generates a Monte Carlo approximation to the non-parametric Maximum Likelihood Estimate of the statistic of interest (p. 33). I will first discuss how the bootstrap method works, and then show how I use it in the context of displaying sets of F1-F2 measurements.

The bootstrap resampling techniques works as follows: We are given a sample $S_1$ of $N$ observations. Create a new sample of the same size by randomly choosing $N$ data points from $S_1$, with replacement, where the probability of choosing any data point at any time is constant, at $1/N$. Then calculate the desired statistic using this new sample; this is a single re-estimate of that statistic. Repeat this procedure some large number of times (e.g., 200, 1000), and consider the distribution of the resampled statistics.\footnote{A function in the S language (described in Becker, et al, 1988) to bootstrap resampled statistics from a given dataset is given here and described below:}

The re-estimated statistics fall in a certain distribution, which may be viewed as a histogram. This histogram is an approximate estimate of the sampling distribution of the original statistic, and any function of that sampling distribution may be estimated by

\begin{verbatim}
bootstrap <- function(data,stat,nResamplings) {
    N <- length(data)
    result <- vector(mode="numeric",length=nResamplings)
    for (i in 1:nResamplings) {
        result[i] <- stat(data[round(runif(N,min=-0.5,max=N-0.5))])
    }
    return(result)
}
\end{verbatim}

This function works thus: Inputs are the sample itself, $data$; the function, $stat$, which generates the statistic (e.g., mean, sum, variance, etc.); and the desired number of times to resample the data, $nResamplings$. The length of the $data$ vector is the number of observations in the sample, $N$. The output of the function, $result$, is a vector of numbers as long as the number of resamplings desired (this should be on the order of 200 to 1000). The substance of the function is a single line inside a loop, which repeatedly does the resampling and the calculation of the statistic. The key line works from the inside out, thus:

$\text{runif(n,min,max)}$ generates a list of $n$ random floating-point numbers from a uniform probability distribution between $min$ and $max$, thus generating random floating-point numbers between $-0.5$ and $N-0.5$. $\text{round(vector)}$ rounds off each element of a list (or vector) of numbers to integers; in this context it rounds off all the random numbers to integers between 0 and $N-1$, inclusive. $\text{data/vector}$ uses the vector of integers as a list of indexes into the sample data, and extracts the indexed elements. Thus the random numbers are interpreted as indexes into the data array. Note that multiple references to the same index can occur freely, so that if $data$ was the vector $[6, 7, 8, 9]$, then the expression, $\text{data[[1,2,1,3,1]]}$, would return the vector, $[6, 7, 6, 8, 8, 6]$. $\text{stat(vector)}$ interprets $vector$ as a data sample, and computes the given statistic using the numbers in that sample. In this context, then $\text{stat()}$ takes the indexed data-points picked by the random-number generator, and calculates the statistic using them. Finally, $\text{result[i]} <- \text{stat(\ldots)}$ sets the $i$'th element in the $result$ vector to be the value that is returned from the calculation, $\text{stat(\ldots)}$.

This code is not especially fast, and it is elaborated when applied to multi-dimensional data, but it is effective both for doing the task itself in the one-dimensional case and for showing how to do (and how easy it is to do) bootstrap resampling.
calculating the function on the basis of the histogram. For example, the standard deviation of the re-estimated means is a reliable estimate of the standard deviation of the mean (that is, the average amount of scatter inherent in the statistic, or the average distance away from the mean at which other similarly constructed means would occur). Another application is to estimate confidence intervals for a given statistic (Efron 1982:78,80), though this is less reliable, since the problem is more difficult, requiring more precise estimates of every detail of the distribution.

Bootstrap methods apply just as well to many-sample situations and to a variety of more complicated data structures" (Efron 1982:35). In the present situation, we are interested in estimating the sampling distribution of the mean of the set of F1,F2 measurements taken for a given phonological class. This is a two-sample problem (i.e., the measured data points are two-dimensional), to which the bootstrap applies. The sampling distribution of the mean amounts to all aspects of the distribution of a statistic, to which the bootstrap technique also applies. The result of applying the bootstrap technique is, again, a Monte Carlo approximation to the non-parametric Maximum Likelihood Estimate of the statistic of interest. Here, this means that the distribution of the re-estimated means generated by the bootstrap technique is an approximation of the sampling distribution of the mean.

One way to look at a distribution of bootstrapped re-estimates is to consider that the "true" value of the statistic could with equal probability be any one of the re-estimated values. Thus the distribution of re-estimated statistic shows the range within which the true statistic (given infinite similar data) might fall, and the density of re-estimates at any one point shows how likely the true statistic is to fall at that location. The true location of the statistic (given an infinite sample) is most likely to be located where the re-estimated statistics are most tightly clustered.

The bootstrap method can be used in many ways, but I will use it rather cautiously, simply as a method for displaying estimated means in F1-F2 space. A number of plots are given in later chapters which display the distribution of 200 or more bootstrapped re-estimates of the mean of a cloud of F1,F2 measurements. These bootstrapped mean distributions are to be interpreted as estimating the range within which the true mean (given an infinite sample of similar data) is located. These distributions are small clouds on the page; the true mean is no more precisely located, according to the given sample of
data, than the area covered by that cloud. If two clouds do not overlap, or overlap by less than 5%, say, then the difference between the two categories is statistically significant.

This enables a visual evaluation of the significance of differences among measurements for various phonological categories. If all the clouds on the page are non-overlapping, then they are pairwise statistically significantly different, and the relationships among the clouds provide a rough estimate of the relationships among the true means for the distributions from which the measured data was sampled (more precise estimates are the point-to-point differences among the original means which lie in the centers of each of the clouds).

5.9.2 Are the Means of Two Samples in 2-Space Significantly Different?

Given sample A and sample B, sets of two-dimensional measurements: Are their (two dimensional) means (or “centroids”) genuinely different? To take a slightly more concrete case, Is the effect of a linguistic factor on F1-F2 measurements significantly at a 5% level of confidence? The statistical problem is that of numerically estimating the significance of the difference between means of two (or possibly more) bivariate samples (which cannot be assumed to have similar (co)variances).

The method used here was the following. Consider the line between the means of the two sets A and B. Project all the points from both sets onto that line (that is, find the point on the line from which a perpendicular will intersect the datum). Thus the data are reduced to one dimension, namely, distance along that line from some arbitrary reference point. The standard test for statistically distinguishing two one-dimensional data samples, namely the t-test, is applied. Since the t-test assumes equal variances between the two sets, this assumption is initially tested using the usual F-ratio test. If the variances are significantly different, then the unequal-variances t-test is used instead of the standard t-test. These tests are discussed in Press, et al, (1988, Numerical Recipes in C, Chapter 13.) Software in the S data analysis language was written to compute these statistics.

All the information about the differences between two distributions is not retained when the two-dimensional data are reduced to one dimension. Nonetheless if this test indicates that the difference between the means is significant, then there is a significant difference
between the two sets of measurements in the direction of the line between the two means.
There may also be other differences as well, which are not pointed out by this test.
Chapter 6

Jamaican Creole

This chapter examines the phonetics and phonology of vowels in Jamaican Creole (JC). JC is quite different from the other English dialects studied in this thesis. When untrained Americans listen to the speech examined here, they understand very little of it, if any. Nonetheless JC is historically related to the American dialects. It is the descendant of a 17th century creolization process which, simply put, consisted of West and Central Africans acquiring and nativizing the vernacular and dialectal British Englishes (including significant exposure to Irish and Scottish varieties) which their forced labor brought them in contact with. This chapter is a purely synchronic analysis. Historical issues are discussed separately in Appendix 2, which presents some evidence for chain shifting in Caribbean Creoles, and discusses the important role of the principle that mergers are irreversible in language history, in “decreolizing” creoles in general, and in Jamaican Creole in particular.

This chapter follows the general pattern of the other chapters on individual dialects. I begin with some background information about the speakers. I then discuss the surface phonological inventory and structure of the vowel system, some impressionistic transcriptions of the stressed vowels, the overall shape of vowel space, the sound-shifts that are ongoing or completed in this dialect, and the effects of stress on vowel quality. To my knowledge this is the first acoustical study of JC vowels.

A phonetic grammar is presented which derives the relationships among of mean locations of vowel classes in F1-F2 space. This grammar applies general principles of phonetics and of sound change in a language-specific way to specify the phonetic quality of vowel
nuclei. A small study of phonetic vowel length was also conducted.

6.1 Speaker Information

The speech studied here is not the deepest creole found in the countryside and mountainous regions of Jamaica, nor is it the upper-class, British R.P.-influenced "acrolect" of Jamaican society, which is more understandable to outsiders. The speech data examined in this chapter comes from a corpus of more than 60 interviews made by Peter Patrick as part of a sociolinguistic study in of a neighborhood in Kingston, Jamaica. This corpus forms the best available data on middle- and lower-class urban speech forms in Jamaican Creole. The speech studied here is quite far from the speech studied in laboratories or with word-lists and self-conscious examination of intuitions. This speech is natural, unmonitored, informal, and vernacular: that is, it is the normal form of language.

The primary speaker studied, Juba, is a 28-year-old, male, working-class, urban-oriented speaker of mesolectal Jamaican Creole. Juba's social background and speech are typical of a mesolectal Jamaican Creole speaker, partly of the country and partly of the city. He was born in the capital, Kingston, and like many children born in the city, was raised in the countryside until the age of 14, when he returned to Kingston to stay.

Juba works as a salesman and lives in Veetan, the mixed-class East Kingston neighborhood where Patrick did his work. A female friend was also present during the interview, which took place in a parked car on a quiet lane. As is also typical of younger, male, working-class speakers in the United States, Juba was interested in telling the fight stories that formed the data for this study.¹ Fights are socially defining events, and stories about fights are very useful in learning about what is considered important to the speakers, as well as exciting stories in themselves. Juba was quite involved in telling the three narratives about childhood fights that constitute his data. Thus his speech was relatively un-self-conscious, and the speech studied here is in a relatively vernacular speech style.

The second speaker analyzed, Roasta,² is an upper-working class speaker. He was raised in Mandeville, one of the larger towns on the island (pop. 40,000) and attended a

¹ As was Jim from Chicago, for example.
² Unstressed /ə/ becomes /a/ in Jamaican, thus I might call him "Roaster", a nickname that comes from his avocation, "roasting", or moonlighting with his employer's equipment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
technical school there, completing 11 years of school. This puts Roasta in the top 10% of the Jamaican population in terms of educational achievement, according to the 1982 census. He lived on the edge of Veeton for nine years after school. Though recently, after getting married, he moved out to a better neighborhood, his primary friendship network is still in Veeton. Roasta is 29 years old, married, has a two-year-old daughter, and makes a good income as a tool-and-die man and roaster, relative to the unskilled workers, small shopkeepers, and clerks in the neighborhood Patrick studied. His urban, as opposed to rural, orientation is characteristic of young people in modern Jamaica, where urbanization is a powerful force behind the social changes sweeping the country. Roasta was interviewed at dinner with his best friend from school days, also a former Veeton resident; their discussion about their former girlfriends is relaxed, interesting, and occasionally quite animated; the presence of his friend, the food, and the topic of discussion all contribute to making this one of Patrick’s best (least monitored) interviews.

6.2 Surface Phonological Structure

The phonology of Jamaican Creole vowels has been studied by several, including Cassidy, LePage, DeCamp, and Wells. My main source for the following phonological analysis is Wells, who included a short summary of JC phonology in his 3-volume Accents of English (1982). Let us review the set of phonological contrasts. Jamaican Creole English does not distinguish the following groups of lexical sets, according the description in Wells (1982):

The phonological inventory (of stressed vowels) may be structured using 3 heights, backness, and length, plus 5 underlying diphthongs, as in Table 6.1 The symbols are Cassidy’s.

A distinctive feature of the Jamaican vowel system is the relationship between Middle

---

3This information is from the Statistical Yearbook of Jamaica, 1986.
5As discussed in Phonological Preliminaries, page 42, Wells’ comparative categories are not the same as the historical word classes whose mergers, splits, and phonetic changes resulted in the modern form of the language. These categories are not primarily intended to show the changes by which Middle English developed into Early Modern English and into Modern Jamaican, but simply to show the lexical correspondences that now hold between this dialect and others.
†The uncertain, possibly unmerged status of the low back phonemes /ɔ, ɔː/, which extensionally correspond to the lexical sets LOT, THOUGHt, NORTH, is discussed below and in Appendix 2.
6The corresponding lexical sets are shown in a footnote on page 164.
Figure 6.1: Lexical sets that are not distinguished in Jamaican Creole

/a/ LETTER = COMMA (unstressed)
/a/ TRAP = LOT'
/a:(r)/ BATH = PALM = CLOTH' = THOUGHT' (= START = NORTH')
/ai/ PRICE = CHOICE
/ir/ NEAR = SQUARE
/ur/ CURE = FORCE

Table 6.1: Jamaican Creole Vowel Structure (Base-6)

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>V:</th>
<th>Vr</th>
<th>Vi</th>
<th>Vu</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i</td>
<td>u</td>
<td>ir</td>
<td>ur</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>uu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>o</td>
<td>ie</td>
<td>uo</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>a</td>
<td>o'</td>
<td>aa</td>
<td>o:'</td>
<td>ar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

English //i:://, //u:://, (corresponding generally to the lexical sets PRICE and MOUTH, Reference American /ay, aw/, and JC /ai, ou/). In some English dialects the nuclei of these vowels have fallen together, but in Jamaican, the nucleus of ME //u::// has not fallen as to low position, and these nuclei remain distinct, as shown by the phonetic measurements on page 172, and as symbolized in Table 6.1 by the mid-back form, /ou/.

It should be pointed out that the apparent lack of distinction between the lexical sets PRICE and CHOICE is false, since the CHOICE class in some cases has an onglide /w/, so that by and boy, for example, are distinguished as /bai/ and /bwai/. However I follows the traditional analysis in which the nuclei and onglides of the two sound classes are considered to be the same, and the distinction between them is no longer in the rhyme of the syllable.

JC /o/, which corresponds to Reference American /ʌ/ (lexical set STRUT), is a mid-back short vowel. This phonetic form is appropriate to the phonological features assigned to this vowel category by the analysis in Phonological Preliminaries. There, /ʌ/ was put into the non-front, mid, short slot; it is phonetically central rather than back in RA. In

---

7Although avoid is /avaid/ rather than /avwaid/ in Roasta's speech, so merger has indeed occurred in some cases.
JC, the corresponding phoneme is a back vowel, which may be taken as support for the earlier analysis.

The structural analysis represented by this chart is simpler and more abstract than the traditional analysis of Cassidy (1907), in which more post-nuclear glides are posited. He transcribes the long-mid vowels in a way close to the phonetic surface (for stressed vowels), namely as /ie, uo/ (rather than /ee, oo/, or /e:, o:/ for example, as per Wells), which are high vowels with downward glides (that is, inglides). The charts below use Cassidy’s transcription. In Table 6.1, these vowels are located in the gaps in the long-vowel subsystem, and thus may be more abstractly written as /e:, o:/ The more abstract system can generate Cassidy’s system by an application of V/-V Raising, which raises the nuclei (but not the glides) of the long mid vowels, as discussed below.

The long vowels, both mid and high, are phonetically raised and shifted to the periphery relative to the corresponding short vowels. (This is shown by the mean phonetic locations of these classes in Figures 6.6 and 6.7, pages 174 and 175.) /uu/ is more backed than raised relative to /u/, while /ii/ is more raised than fronted relative to /i/, but this basic generalization still applies to the high vowels, as well as to the mid vowels. This would appear to be another instance of Vowel-before-Vowel Raising.

The mid vowel, /e:/ (here labelled “ie”), is raised as high as /ii/ for Roasta and nearly as high as /ii/ for Juba. This justifies the “i” in Cassidy’s “ie” transcription. /o:/ is labelled “uo” for the same reason: its nucleus is raised towards [u], and it has a lowering offglide.8 The presence of the offglide and the raised nucleus with the long, phonologically mid vowels provides an opportunity to apply the phonological representations developed in Phonological Preliminaries. If these long mid vowels are underlyingly represented in the same way as the long mid vowels of Reference American, then a plausible derivation from underlying to surface form is in Figure 6.2.

The phonological analysis in Figure 6.2 raises two important technical issues. First, notice that the feature [hi(gh)] after step 3 is not really [high]; rule (3) merely adds a degree of phonetic height, which is not sufficient to reach the full phonetic height of the

---

8The offglide could be [9] or [Ã], [°]. Distinguishing these possibilities is a difficult matter, since [°], [u°], and [°, u°] are all ingliding vowels. It may have been equally appropriate if Cassidy had written these vowels as /ia, u3/.
Figure 6.2: Derivation of ingliding for long mid vowels.

\[
\begin{array}{ccccccc}
  & N & G & N & G & N & G \\
\mid & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\
\mid & \text{-hi-} & \text{-lo} & 1 & \text{-hi-} & \text{-lo} & 2 & \text{-hi-} & \text{-lo} & 3 & \text{hi} & \text{-hi-} & \text{-lo} \\
\end{array}
\]

1: glide-feature-specification.

2: separation of glide and nucleus feature-specifications

3: vowel-before-vowel raising of the mid nuclei

long high vowels, particularly in the case of /uo/, as shown in Figures 6.6, 6.7, pages 174, page6.7. The raising is not really the addition of the feature [high] to a discrete feature specification, since phonologically [high] vowels are actually even more raised. So the issue is, What categories does this raising rule operate over?

There appears to be a gradient dimension of phonetic height, since there are a fairly large number of different heights distinguishing raised and non-raised vowels. The back vowels may be arranged in a scale from high to low according to the mean F2 for stressed tokens of the vowel, thus: /uu, uo, ou, o, aa/. The front vowels may be arranged in a similar scale, /ii, ie, i, e, a, a/. The five or so heights phonetically necessary here may be distinguished by using a small-n numerical system of height specification such as Clements', Schane's, or Ladefoged's (see Phonological Preliminaries), or may be represented by a more continuous gradient, as assumed in Labov, Yaeger and Steiner (1972). Since the phonetic height represented in the chart of acoustic measurements must be derived at some point, we may economize on intermediate steps in the derivation by considering this a phonetic raising rule that operates on a continuous gradient. The continuous gradient must be brought in at some point anyway, since it is necessary to account for the acoustic observations and the articulatory differences that they reflect. If we account for the phonetic raising of long non-low vowels by applying the rule used in step (3) above at this low level in the derivation of phonetic forms, we can avoid postulating an additional, unnecessary level of phonetic derivation.

In fact this general line of argument suggests that small-n numerical height specifications are to be avoided if possible, in favor of phonetic rules that interpret phonological categories in terms of continuous, gradient phonetic dimensions. The precise form of such
rules is yet to be ascertained, but in this case, their effect is clear: the nuclei of the long, mid vowels are raised to some degree; the nuclei of short vowels are lowered by some scalar value.

A second technical question is, how is the natural class of mid vowels to be specified in the specification of rule (3)? They are neither high nor low, so we may say positively that the height tier\(^9\) lacks specification, and thereby pick out the mid vowels. Since this raising rule applies to both front and back mid vowels, I don’t specify the [front] feature. But the lack of specification may have two quite different meanings, as discussed in Chapter 3, page 77. The lack of specification of the [front] feature is used to avoid all values on a particular dimension, namely both [front] and non-[front]. The lack of specification of [high] or [low], on the other hand, is used to pick out a particular point on the height dimension, namely the height which is neither [high] nor [low]. There is no formal difference between not specifying a feature in order to generalize over the values of the feature, and not specifying a feature in order to pick out the unmarked value on that dimension. This is a general problem with privative features in phonology.

In the current case, we’re lucky, and there is no conflict between these two significances, but such occasions could conceivably arise, posing severe difficulties to theories relying on underspecification, without some solution like that given in Chapter 3.

Despite the improvements in structural symmetry and simplicity, and the elimination of gaps gained by analysing Cassidy’s /ie, uo/ vowels as long mid vowels, the base-6 system represented above has two major faults: it makes a fundamental distinction between low-front and low-back vowels that may be unnecessary, and it still contains a great many gaps among the gliding vowel subsystems.

As in the phonology of Reference American vowels, some of the central problems in characterizing Jamaican Creole vowel structure are the low vowels and the glides. The various low vowels classes of other dialects are collapsed in JC into just two: long and short a, variously written as /aa, a/ or /a:, a/. The lexical sets which make up these phonemes are listed above in Figure 6.1. However, this picture holds two mysteries: the low-back vowels, and /ar/.

---

\(^9\)In Phonological Preliminaries, phonological height is analysed as an autosegmental tier, which may contain a single privative height feature or none, where [high] and [low] are the two features available.
The low-back vowels, marked with † in Table 6.1 and Figure 6.1 above, have an uncertain phonological and historical status. The phonemes /a, a:/, when they occur in words of the historically back-round lexical sets, LOT, THOUGHT, CLOTH, and NORTH, are said to be optionally backed and rounded in acrolectal Jamaican Creole English. However, this backing is said to be impossible for /a, a:/ in words which historically were not back and round, namely PALM, BATH, and TRAP words. This poses a paradox, since this situation is not compatible with the following widely accepted assumptions:

1: Low-back and low-front vowels are merged in the basilectal creole.

2: The more acrolectal (RP-like) levels of the creole are historically descended from the earlier basilect through a process of decreolization.

3: Merger is irreversible.

Given these conditions, it is impossible that the merged front and back low vowels should somehow be unmerged in the acro- and mesolects of JC. These assumptions are discussed extensively in Appendix 2. I will assume, for now, that the accepted assumptions 1, 2, and 3 are indeed true, and instead that the claim that /ɔ, ɔ:/ are phonologically distinct from /a, a:/ is false.

The status of the phoneme /ar/ (with the START and NORTH lexical sets) is also something of a mystery. This vowel is long and has the same nucleus as /a:/; Commonly there is no rhotic glide pronounced with these words, so that it seems that they should be classified with the /a:/ phoneme. However, words in the START and NORTH classes are occasionally pronounced with a rhotic glide, and the other classes are almost never pronounced with a rhotic glide (Patrick 1991). Thus it would appear that /ar/ is a separate phonological category, which is frequently neutralized with /a:/.

A well-designed phonological system, both from the perspective of the linguistic analyst, and from the perspective of the living, changing language itself, is symmetrical and gap free. Gaps in phonological structures count as flaws in two senses. First, they may be real and then often result in historical splits, mergers, or sound shifts, by which the gaps are eliminated. Second, gaps may be due to faulty linguistic analysis. If the system is historically stable, as Jamaican seems to be, then attention may be turned to the analysis.
Notice that there are three kind of glides (/i, u, r/), but only 5 gliding vowel phonemes (excluding /ie, uo/). The situation is somewhat similar to the table from Trager & Bloch (1942), discussed in Phonological Preliminaries, where gaps and fillers correspond with each other across subsystems. We might reduce the Nucleus-Glide sequences from two or three paradigms with as many as 13 gaps to just one, with no gaps, thus:

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>ir</td>
<td>ur</td>
</tr>
<tr>
<td>Mid</td>
<td>ai</td>
<td>ou</td>
</tr>
<tr>
<td>Low</td>
<td>ar</td>
<td></td>
</tr>
</tbody>
</table>

This rearrangement considerably simplifies the phonological structure at the cost of additional rules. The following four rules will do the necessary rearrangements. They use binary features [±high], [±low], [±rhotic], and [±front], labelled with “N” for features that occur in the Nucleus slot in the syllable, and with “G” for features in the Glide position. Here I assume Glide segments in the N+Glide subsystem are specified as [-high] at the point in the derivation where these rules apply.

1. \( G[-\text{high}] \rightarrow G[+\text{high}] /N[-\text{high},-\text{low}] \).
2. \( G[-\text{high}] \rightarrow G[\text{rhotic}] \).
3. \( G[+\text{high}] \rightarrow [\alpha \text{ front}] /N[\alpha \text{ front}] \).
4. \( N[+\text{front}] \rightarrow [+\text{low}] /\_\_G[+\text{high}] \).

Informally, these rules arrange it so that the non-mid vowels have rhotic glides, while the mid vowels have high glides with a frontness feature inherited from the nucleus. The Glide slot for this set of vowels underlingly contains the feature [-high], which is the underspecified form of the [rhotic] glide, as argued in Phonological Preliminaries. [-high] is rewritten as [+high] before mid vowels in (1), which is thus a rule of nucleus-glide differentiation. Then the remaining [-high] glides are rhoticized by rule (2), which must follow (1) as stated here. (3) assimilates the non-rhotic, high glides to the frontness of the adjacent nuclei, and must also follow (1).
The last rule, (4), changes the nucleus of the mid, front, upgliding vowel to low. This rule recapitulates a late part of the Great Vowel Shift,\textsuperscript{10} lowering the mid nucleus of a rising diphthong. It is restricted to apply only to the front mid vowel and not the back.\textsuperscript{11}

We have hereby eliminated two underlying vowel-pause paradigms or subsystems, as well as the low-back vowels, from the base-6 system represented above in Table 6.1. The result of these eliminations is a rather abstract, but extremely simple and symmetric base-5 system of short, long, and gliding vowels, represented in Table 6.2.\textsuperscript{12}

Table 6.2: Jamaican Creole Vowel Structure (Base-5)

<table>
<thead>
<tr>
<th>V</th>
<th>V:</th>
<th>V-Glide</th>
</tr>
</thead>
<tbody>
<tr>
<td>high:</td>
<td>i, u</td>
<td>i: u:</td>
</tr>
<tr>
<td>mid:</td>
<td>e, o</td>
<td>e: o:</td>
</tr>
<tr>
<td>low:</td>
<td>a, a:</td>
<td></td>
</tr>
</tbody>
</table>

It is remarkable that the analysis of Reference American given above in Phonological Preliminaries resulted in a structure quite similar in fundamental respects to that argued for here: 3 heights vs. backness vs. length, plus restricted combinations of nuclei with 3 glides. Despite the important phonetic differences and the major differences of inventory, the structure of the vowel system of Jamaican Creole described here has the same basic dimensions as the surface phonological structure of a radically different dialect of English. The differences are that this representation of the Jamaican vowel system has a base-5 rather than a base-6 structure (that is, there is only one low vowel per subsystem), and that the gliding vowels can be collapsed into a single subsystem.

\textsuperscript{10}If we may reason from synchronic rules to diachronic rules, it would appear that the Great Vowel Shift did not proceed symmetrically in the front and the back. It is indeed a fact that the two vowels that underwent diphthongization and lowering from earlier //i, u// to modern /ay, aw/ in many dialects of English did not both fall to low in Jamaican Creole.

\textsuperscript{11}The set of rules stated here could possibly be simplified if the natural rules of nucleus-pause differentiation (in height features) and nucleus-gliding assimilation (in backness features) were attributed to general principles, and if they did not need to be stated explicitly in a grammar of Jamaican Creole.

\textsuperscript{12}The lexical sets corresponding to the slots in Table 6.2 are as follows:

<table>
<thead>
<tr>
<th>V</th>
<th>V:</th>
<th>V+ (i, u, r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT/FOOT</td>
<td>FLEEC/E</td>
<td>GOOSE/N</td>
</tr>
<tr>
<td>DRESS/STRUT</td>
<td>FACE/GOAT</td>
<td>NEAR/SQ</td>
</tr>
<tr>
<td>TRAP/LOT</td>
<td>PALM/THOUGHT/CLOTH</td>
<td>PRICE/CHO</td>
</tr>
</tbody>
</table>
6.3 The Shape of Formant Space

2680 and 1673 vowels were measured for the two speakers, according to the methods described in Chapter 5; outlyng measurements were examined individually and either corrected or confirmed. As argued in the Acoustics chapter, measurements of the first two formant frequencies reflect the articulatory dimensions of mouth-opening and of tongue-body backness and lip rounding. All the F1, F2 measurements for the two speakers are displayed in Figures 6.3 and 6.4.

Comparing the overall shape of the vowel space (that is, the distribution of measurements of F1, F2 tokens) across dialects shows some real differences in the envelope of acoustic variation (Labov 1991).

These vowel spaces differ from those of other dialects in that the location of maximum density of tokens — that is, the mode of the distribution — is closer to the bottom corner of the triangle. The Chicano vowel space has two modes in the high front and high back; while the Chicagoans’ vowel spaces are more evenly distributed. None of the other speakers, including the Alabama speaker, has a mode in the low corner. Since both Juba and Roasta have modes in the low corner, this feature is not an individual idiosyncrasy.

Juba’s vowel space, which was analysed most extensively, is quite different from those found for the other English dialects (pages 198ff, 233, 254). The shape of this vowel space is roughly triangular, or rather, V-shaped, with tokens relatively sparsely distributed at the top, and in the middle of the V. The paucity of tokens in the upper-middle region of the distribution is consistent with the claim that the high-central region of vowel space is typically unoccupied (cf. Liljencrantz & Lindblom, 1972). The distribution of tokens is densest on the front and back edges of the triangle, and more dense towards the bottom. Given that the overall shape is a triangle, two rules summarize the picture: the lower, the denser; the closer to the front or back edge, the denser. This makes the distribution look like a bottom-heavy V shape.

As with all speakers studied in this thesis, there is some front-back asymmetry in the overall envelope of variation. The distributions are rather more dense in the high-front than in the high-back, and the back edge of the triangle is possibly more vertical than the

---

13The numbers of measurements broken down by class and speaker are given in Appendix 3, along with mean vowel durations.
Figure 6.3: 2680 F1/F2 measurements.

Jamaican Vowel Space: Juba, 28

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 6.4: 1673 F1/F2 measurements.

Jamaican Vowel Space: Roasta, 29

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
front edge (a difference which may be eliminated by transforming the scale to barks, mels, or a logarithmic scale.)

6.4 Impressions of Stressed Vowels

The first ten stressed tokens of each vowel in Juba’s data were transcribed impressionistically, to give an idea of the allophonic range of the vowel in stressed environments. The range of transcriptions are as follows:

/ii or i:/  [i, i:, i', i:]  /i/  [i, i, i, i', i]
/ie or e:/  [i, i, i, i, e, i, e, i]  /e/  [e, i, e]
/aa or a:/  [a, a, a]  /a/  [a, a, a, a]
/uo or o:/  [u, u]  /o/  [o, o, o, u]
/uu or u:/  [u, u, u, u, u, u]  /u/  [u, u]
/ai/  [a, a, a]  /ou/  [a, a, a, o, o]
/ir/  [i, i, i, i, i, i, i, i]  /ur/  [u, u, u, u, u, u]
/ar/  [a]  /ar/  [a, a, a]

Note that the long “mid” vowels are never as low as mid when stressed. The long vowels surface as monophthongs more than in RA. The long-short distinction is much more important on the phonetic surface; the phonetic glides which are intrinsic to the underlyingly long vowels in RA occur less commonly in JC: monophthongs are quite common on the surface for the long vowels /ii, aa, uu/. Here phonetic length and quality rather than phonetic gliding is the key to the distinction between the long and short vowels. These differences are explored instrumentally in studies of vowel duration and of the phonetic quality of the vowel nuclei.

6.5 Acoustic Correlates of Phonological Vowel Length

One of the main phonetic differences between JC and other English dialects is that length is more clearly realized as such in Jamaican than elsewhere. This section will document the phonetic implementation of phonological vowel length in JC.
6.5.1 Duration Differences.

Appendix 3 shows mean acoustic vowel durations for each vowel phoneme, for stressed and unstressed vowels, and for vowels before and after voiced and voiceless consonants, for each speaker. Standard deviations and standard errors are also given there. Note that the means given there are grand means; some of them may be skewed by uncontrolled interactions with other effects. Many interactions will be washed out by the law of large numbers, but perhaps not all, so some caution must be used in interpreting the numbers.

Table 6.3 was constructed from the data in Appendix 3. For each short/long pair (that is, /i, ii/, /e, ie/, /a, aa/, /o, oo/, /u, uu/), the mean vowel durations are presented, and the ratio of the mean durations of short and long vowels is expressed as a percentage. The mean durations for all short vowels together and all long vowels together (including diphthongs, in this case) are also given, along with the corresponding ratio.

Table 6.3: Jamaican mean vowel durations (ms) and short/long duration ratios (percent). Diphthongs (/ai/, /ou/, etc.) are included in “all” long vowels.

<table>
<thead>
<tr>
<th></th>
<th>Juba</th>
<th>Roasta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V:</td>
<td>V</td>
</tr>
<tr>
<td>/i:/</td>
<td>116</td>
<td>/i/</td>
</tr>
<tr>
<td>/e:/</td>
<td>135</td>
<td>/e/</td>
</tr>
<tr>
<td>/a:/</td>
<td>190</td>
<td>/a/</td>
</tr>
<tr>
<td>/o:/</td>
<td>122</td>
<td>/o/</td>
</tr>
<tr>
<td>/u:/</td>
<td>133</td>
<td>/u/</td>
</tr>
<tr>
<td>all</td>
<td>144</td>
<td>all</td>
</tr>
</tbody>
</table>

For the two Jamaican speakers studied, phonological vowel length (including diphthongs as long vowels) approximately doubles the vowel duration in the aggregate (64:125ms for Roasta, 76:144ms for Juba). Roasta generally spoke more rapidly than Juba, which accounts for the generally lower segmental durations measured for his speech.

If slower speech enables underlying durational patterns to manifest more clearly, then this would account for the smaller ratios (= greater differences) between Juba’s long and short vowels than between Roasta’s long and short vowels.

The details of the long/short relationship differ in interesting ways for different vowels. In particular, the mid vowels stand out from the others, for both speakers, as seen in the

169

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ratios of mean durations in Table 6.3, expressed as percentages. These ratios are among
the main phonetic effects of phonological vowel length. If average duration of acoustic
vowels corresponding to a long vowel were twice the average duration of acoustic vowels
corresponding to short vowels, then this ratio would be 50%. Here we see that the high
and low vowels have similar ratios for each speaker, while the mid vowels have considerably
higher ratios. That is, the durational difference between long and short vowels is relatively
smaller for the mid vowels, because the ratios are closer to 1, or 100%.

Why should phonological length correspond to a smaller difference for mid vowels than
for high and low vowels? The long mid vowels are phonologically special in the Jamaican
vowel system: their nuclei are raised (when stressed) and they become phonetically gliding
vowels, so that duration and peripherality are not the only phonetic features distinguishing
them from the corresponding short vowels. The high and low long vowels, on the other
hand, do not become glides; they largely remain monophthongs even when stressed.

These facts may be related. That is, the special status of the mid vowels would seem
to be related to their distinctive patterning with respect to acoustic vowel duration. It
is tempting to speculate that the phonetic process by which the vowel becomes a glide
has a weakening effect on the durational difference between the long and short vowels.
This hypothesis is supported by the fact that in the other dialects studied in this the-
sis, the long vowels generally have phonetic glides, and furthermore phonological vowel
length corresponds to a smaller difference between the vowel durations. Thus, the ra-
tios between the average short vowel duration and the average long vowel duration for
the other speakers studied are 90/126ms=71% (Judy, Chicago), 81/109ms=74% (Jim,
Chicago), 91/107ms=85% (Rita, Chicago), 105/148ms=71% (James H., Alabama), and
79/102ms=77% (Vince, L.A. Chicano). These ratios for Juba and Roasta are 52% and
51% (including diphthongs), showing a much greater difference in phonetic duration, cor-
responding to phonological vowel length in Jamaican Creole.

Thus there may be a general association of phonetic gliding of long vowels with a
weakening of the durational differences between the corresponding long and short vowels.
Long vowels, which have phonetic glides in the other dialects, are phonetically not as
much longer than the corresponding short vowels. The long vowels that in Jamaican have
phonetic glides are also not as much longer, compared with the long monophthongs.

170
Whether or not this general association between phonetic gliding of long vowels and a relatively difference between the durations of long and short vowels is born out in later studies, it is clear that phonological vowel length and gliding in Jamaican Creole is realized by very large durational differences, with ratios in the neighborhood of 1:2.

6.5.2 Phonological Length and Vowel Quality

What are the effects of phonological vowel length on vowel quality? Insofar as formant-frequency measurements reflect perceptible vowel quality, we may explore these effects by plotting such measurements for pairs of long and short vowels. Consider the charts in Figure 6.5, which displays the five short and long vowels, /i(:), e(:), a(:), o(:), u(:)/, one pair per chart, for Juba, the first speaker studied. These charts show all tokens of the given phonemes, including all segmental environments, and all stress levels.

The expected result of examining a large number of tokens of different vowel phonemes, where short and unstressed tokens are included, is that the classes will overlap. There are many contextual effects that modify vowel quality: The vowel in well may sound closer to that in wool than that in head. As Lisker (1948) showed, even measurements of a vocalic minimal pair (pek, pex) in monitored laboratory speech may overlap in F1-F2 space, though this result was not confirmed in a replication conducted as support for the methods used here (page 118). The problem in natural speech is much worse than Lisker's results would suggest: the distribution for a single vowel can cover almost the entire vowel space (as shown on page 202). As Figure 6.5 shows, overlap is no less pervasive in Jamaican than in Chicago. Very little of this overlap may be attributed to errors, as shown in the above discussion of the analysis of outliers (page 146).

Thus we see that /i:/ and /i/ significantly overlap each other, despite having quite different means. Sometimes skim may sound like scheme, but usually not. Similarly for the other long/short pairs: the clouds overlap, but have different means. One exception may be /a:/ and /a/, where the mode of the short /a/ distribution is near the location of the mode of the long /a:/ distribution. The primary difference between /a/ and /a:/ appears to be that the short vowel is more widely distributed, perhaps because as a short vowel, it is more susceptible to context and stress effects.
Figure 6.5: Raw F1-F2 measurements of long and short vowel pairs.

Juba: Long (||) vs Short (_) Vowels

- **ii versus i**
- **uu versus u**
- **io versus e**
- **uo versus o**
- **aa versus a**
- **ai versus ou**
The last chart on the page is of /ai/ and /ou/, which do not form a long/short pair, but conveniently filled the sixth position on the page so that I could repeat the point made above about the Great Vowel Shift. The nuclei of these vowels are identified as the same in some English dialects, such as my own, where they can be transcribed as /ay, aw/. In the Great English Vowel Shift, the nuclei of the high-front and high-back vowels, Middle English long /i:/ and /u:/, dropped down to the bottom center of vowel space, with the result that their nuclei could be identified as the same in some of the daughter languages. The Shift has not had the same result here, since the resulting classes /ai/ and /ou/ have nuclei that are phonetically quite different.

6.6 Mean Vowel Nuclei in JC

Keeping the overlap of the distributions in mind, let's consider how much confidence we may have in the claim that the means are different for these pairs. As discussed in the Methods chapter, I use a rather new statistical technique, called the bootstrap resampling method, to evaluate this kind of claim. This technique is useful in determining the degree of precision to be attributed to any estimated statistic. Here the relevant statistic is the mean of a given vowel's distribution in F1-F2 space. Again, the bootstrap resampling method works as follows. A large number of resampled data sets are constructed by randomly choosing from the original data set (of size N) an equal number (N) of observations, with replacement. The mean of each of the new resampled data sets is calculated and plotted, so that the distribution of resampled means may be seen directly. If less than 5% of the re-estimates for one distribution overlap with those of another one, then they are significantly different with estimated probability >95%. The distribution of resampled means shows directly the amount of scatter intrinsic to the estimate of the mean taken from the original sample.

Making use of this method, Figures 6.6 and 6.7 show the distributions of 200 resampled means for each of the non-rhotic Jamaican Creole vowels, for both Juba and Roasta.

Juba’s chart shows quite clearly that the long vowels occupy the corners of the vowel space, while the short vowels are less peripheral and lower. Each vowel is very clearly different from every other vowel, on average. None of the distributions of re-estimated
Figure 6.6: Juba: Bootstrapped mean distributions.

Juba: 200 Re-estimated means for each vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 6.7: Roasta: Bootstrapped mean distributions.

JC (Roasta): 200 Re-estimated means for each vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
means overlap with any other. In fact, no pair of vowels have any overlap; even /ii, ie/
do not overlap, though they are very close to each other, even more for Roasta than for
Juba. The clear differences here are even clearer when the differences between long and
short vowels is considered. The long and short pairs of vowels have quite clearly different
means. In fact, no long vowel has a nearest neighbor which is its corresponding short
vowel, for either speaker. The converse is true for the short vowels, too. Every vowel's
nearest neighbor is a vowel that is not its short or long counterpart. We can conclude with
complete confidence that long and short vowels have very different average nuclei.

6.6.1 Peripherality

Given that there is an effect of phonological length on (average) vowel quality, how can
we best characterize this effect? In most dialects of English, phonological vowel length
translates into phonetic peripherality (a better-defined term than the more traditional
feature, tenseness — see Labov, Yaeger & Steiner (1972, “LYS”). A peripheral vowel is
farther from the center of vowel space than a non-peripheral vowel.

In Jamaican, though, the only vowels that are not peripheral — that is, occurring on
the edges of the equilateral triangle formed by long /ii, aa, uu/ — are short /i, a, u/. All
the long vowels are peripheral, and even short /e/ is peripheral, since it too is adjacent to
the edges of this triangle.

Comparing the long vowel space with the short vowel space, the clearest effect is that
length is not realized as a “peripheral” location in phonetic vowel space in the simplest
sense, but rather it means, “in the corners” of the vowel space. The short vowels occupy
the center, while the long vowels occupy the corners of the vowel space.

It is again useful to compare this picture with the definition of the feature, peripherality
in LYS (1972:p41ff, p106, etc.). Peripheral vowels are those that are distributed along the
outer edges of a vowel space, while non-peripheral vowels are a step inward from the outer
edges. In the current data, however, the phonetic feature associated with length is not
peripherality in that sense of “distributed along the edges”, but rather a different kind of
peripherality, which may be described as “distributed in the corners”.

The JC picture is somewhat more compatible with LYS' definition when it is combined
with their principles of vowel shifting (p. 106), which state: In chain shifts, peripheral vowels rise, non-peripheral vowels fall, and back vowels move to the front. Long vowels are generally peripheral, so the long vowels should over time end up in the upper corners, while the short vowels should be relatively low. Juba's chart shows just this pattern: the short vowel system is relatively low, the long front-and-back vowels have raised to the corners, and the only vowel that is out of place, /u/, is a back vowel which has fronted. Thus all three of Labov's principles of vowel shifting are supported by this data. We will see below how these principles can be incorporated into a synchronic grammar.

In conclusion, we have seen that the phonological distinction of vowel length in Jamaican Creole is realized by large differences in average duration (long vowels are twice as long as short vowels), and by peripherality differences which have resulted in extremely significant vowel quality differences as shown by distributions in F1-F2 space.

The old story is confirmed: while one may analyse a set of phonological categories as though they are distinguished by a minimal set of structural dimensions, the phonetic realizations of the categories are distinguished by multiple, redundant phonetic differences.

6.6.2 An Acoustic-Phonetic Grammar for JC Nuclei

Figures 6.6, 6.7 show more than merely the presence of differences between long and short vowels; it displays the acoustic structure of the vowel systems of these two speakers, insofar as this structure is shown by F1-F2 measurements at the phonetic nucleus of vowels. On this scale, the long vowels form a large inverted triangle, with the mid-long vowels raised along the front and back edges toward the upper corners. The short vowels form a much smaller pentagon, which is inside and in Juba's case near the bottom of the large triangle.

The relations between the long vowels (the triangle) and the short vowels (the pentagon) may be derived by four simple statements from a reasonable default distribution of vowels in this acoustic space, also characterized by four statements, which may themselves be derived from general facts of vowel phonetics. The first four statements are proposed universal phonetic implementation rules, as presented in Table 6.4.

(1) is a general fact about phonetic vowel space. (2) and (3) follow from the disputed

---

14A display of single means would show the same structure, but would not show how significant the relationships between the means are.
Table 6.4: Universal Phonetic Implementation Rules

(1) The default acoustic vowel space (in the dimensions -F2, -F1) is an inverted triangle.

In a 5-vowel system:
(2) Three of the vowels should fall at the corners, and
(3) the other two should fall in the middle of the front and back edges.
(4) Long vowels are peripheral to short vowels.

claims of Liljencrantz and Lindblom (1972, also cf. Lindblom, 1978) on maximal dispersion of vowel phonemes.\(^{15}\) (4) is a general fact about the relation between peripherality and vowel length. These four statements of general phonetics, in combination with the phonological structure that distinguishes 5 nuclei and length, generate a picture of a 10-vowel acoustic vowel system like the first picture in Figure 6.8.

To generate the second picture from the first, four linguistic phonetic statements are needed, as in Table 6.5.

The procedure in Table 6.5 generates the acoustic distribution of the long and short vowels seen in both the displayed schematic chart (Figure 6.8) and Juba’s actual chart of re-estimated means (Figure 6.6). Roasta’s pattern is only slightly different: the back long vowel is raised even less than Juba’s; the short vowels are not apparently lowered (rule 7 is omitted), and /a/ is slightly to the front of /aa/.

Three of these four statements are closely related to the generalizations about historical chain shifts in Labov, Yaeger and Steiner (1972) discussed above: In chain shifts, peripheral

---

\(^{15}\)Liljencrantz and Lindblom’s theory of maximal dispersion may be criticized in two ways. Labov 1982 has pointed out that their results are based on phonemic data, not the kind of phonetic data which would be necessary to support their claims. If i, a, and u are used in the phonological transcription of most of the languages of the world, it cannot be inferred that most of the languages of the world have vowel phonemes with the phonetic qualities of IPA [i, a, u]. A language might well have, for example, [i\(^1\), v, u\(^3\)] as the main allophones for /i, a, u/. Any three-vowel system is likely to be transcribed with these symbols, no matter what the phonetic targets are. Bessell’s (1991 and forthcoming) studies of the phonetics of vowels in Interior Salish languages make this quite evident. The fundamental distinction between phonetics and phonology is ignored when this inference is made from phonology to phonetics.

L&L’s functional theory suggests that sounds should be maximally far apart, but a principle of minimal effort is also necessary, because sounds are not always maximally separated. For example, languages with just two tones often use the minimum degree of phonetic difference to carry the contrast. Thus there is a balance between maximal distinction and minimal effort. But this amounts to saying simply that sounds are more or less distinct, which is vacuous.

Something must be salvaged from L&L’s theory in order to use it in deriving (2) and (3).

178
Table 6.5: Linguistic Phonetic Implementation Rules for Jamaican Creole

5. Raise the mid-long vowels up towards the corners, the front one more than the back one.
6. Shift the mid-short vowels towards the periphery, thus bowing out the edges of the inner triangle.
7. Lower the short vowels as far as possible, while retaining their front-back relations.
8. Front short /u/ so that it is directly above the bottom corner.
(that is, long) vowels raise (as also in rule 5), non-peripheral (that is, short) vowels lower (as in rule 7), and back vowels move to the front (as in rule 8). Rules (5) and (6) may be related by a drag chain shift: When the long mid vowels are raised, the short mid vowels fall into the gap left behind, becoming peripheral vowels. These linguistic-phonetic rules are intended strictly as statements of synchronic relationships; they determine the locations within phonetic vowel space of the average nuclei of the short and long vowels in Jamaican Creole. LYS' rules of chain shifting, on the other hand, describe diachronic relationships. Since synchronic rules often recapitulate historical changes, it may be that these rules also describe actual historical changes in the development of this dialect. However, they are used here synchronically, as part of the derivation from phonological structure to aspects of phonetic implementation. The fact that the rules also govern whole classes of historical changes confirms their naturalness.

This derivation of the locations of the vowels in the acoustic vowel space is systematic and entirely plausible; further the number of statements it requires is significantly fewer than the number of statements required to simply list the mean first and second formant frequencies for each vowel class (2 formants * 5 vowels * 2 lengths = 20 specifications).

The two remaining vowels must also be located in this space, the raising diphthongs /ai, ou/, which are the only vowels peripheral to the long-vowel triangle. These follow a pattern opposite to that of the mid vowels, where the front vowel is higher than the back one: the front vowel, /ai/, is lower than the back vowel, /ou/.

The above statements characterize the locations of vowel nuclei in formant space, an important part of phonetic grammar. As argued in the Acoustics chapter, F1-F2 space directly reflects the continuous phonetic dimensions of height and backness, which are articulatorily realized by the degree of mouth opening and of tongue-body-frontness and lip-aperture. Thus this is not merely an acoustic description, but a phonetic description, which shows much of the structure of the phonetic grammar of the speaker. The rules relate phonological categories to mean locations of vowel nuclei in this phonetic space, accounting for rather complex details of phonetic distribution in a natural, simple, and general way. The rules largely refer to classes of vowels rather than single phonemes. Also, they are mostly independently motivated in previous general discussions of sound change, phonetic typology, etc.
The status of these rules in linguistic theory is of some interest. Statements (1-4) above appear to be consequences of universal principles of general phonetics\textsuperscript{16} that may not need to be stated in a purely linguistic description. These are similar to the phonetic implementation rules given for Reference American (Section 3.6). Statements (5-8), on the other hand, must be included in the phonetic grammar, since counterexamples exist. Thus, long mid vowels need not be phonetically raised towards the long high vowels (statement 5). See, for example, /ey, ow/ in Alabama, page 236. The short mid vowels need not be peripheral (statement 6), and the short high-back vowel need not be fronted (statement 8). See for example, Jim from Chicago, page 206, where /ε, η/ are non-peripheral, and /u/ is phonetically well to the back of /a/, as far back as /uw/. Since languages may or may not implement these statements, they should be specified in the grammar.

This phonetic grammar is incomplete: Further details must be specified to generate vowel trajectories from phonological specifications, including phonological and phonetic diphthongization, consonant-coarticulation effects, etc. Effects of F3, of formant amplitude, of temporal variation are not dealt with here. However, within the limits of single time-slice measurements of F1 and F2 in characterizing vowel quality — limits within which much of what is linguistically significant about vowel quality can be characterized — the 8-rule grammar presented above accurately describes the mapping from phonological to phonetic structure.

\section*{6.7 Stress Effects}

While there appear to be factors other than stress involved in Jamaican prosody (which may be more like tonal phenomena), there are certainly many clear cases of prominent as opposed to non-prominent syllables. Therefore, despite the unresolved phonology of Jamaican prosody, I impressionistically coded this stretch of speech for phrasal stress, and in this section present the correlation of these classifications with formant-frequency variation.

The method for coding stress is discussed in Section 5.4. After considerable time listening to the rhythm of speech of the two speakers studied, my intuitions about the

\textsuperscript{16}For the distinction between general and linguistic phonetics, see page 104.
location of stress seemed fairly reliable. While for reasons discussed below these developed intuitions are in some cases quite different from the naive intuitions of American listeners, they generally coincided with those of Peter Patrick in the joint coding we did.

A self-consistency test was conducted on 231 syllables, which I classified for three levels of stress on two occasions, 9 months apart. The results are presented in Table 6.6.

Table 6.6: Confusion matrix of phrasal stress classifications on two occasions.

<table>
<thead>
<tr>
<th></th>
<th>October, 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Primary</td>
</tr>
<tr>
<td>July, 1991</td>
<td>55</td>
</tr>
<tr>
<td>Unstressed</td>
<td>5</td>
</tr>
</tbody>
</table>

The frequency of consistent classifications is 202/231 = 87%. If secondary stress is collapsed with either unstressed or stressed, the consistency level increases to 213/233 = 92.2%. This is almost identical to the levels of consistency found between another coder and myself in the same task on a Chicago speaker, discussed in the Methods chapter, page 141.

It wasn’t difficult to apply this procedure, which was previously used with the other dialects studied, to Jamaican Creole.

In JC, prominence seems cued sometimes by an apparent pause before the stressed syllable. Pitch rise is not the only way to achieve pitch prominence: Pitch fall is also prominent. But my impression is that this is less frequent.

There are some problems with stress coding in this way. In some examples, there appears to be an unlikely stress clash: three or four consecutive syllables can have high pitch, each apparently stressed. Also, the location of the stress is sometimes unclear, because it may not make linguistic sense to stress closed class words, or because a single pitch prominence may seem to extend over more than one syllable. In fact, Wells (1973:22, 1982:573) says that stress may appear to shift to the right one syllable, from the expected (Received Pronunciation, presumably) stress pattern. This is consistent with the view that stress is a feature on an autosegmental tier, which can slide around from one syllable to the next or previous, or to stretch out across several syllables. This view is rather more
reminiscent of tone than of syllable stress.

Let us consider the effects of impressionistically coded stress on vowel quality as shown in F1-F2 charts of nucleus measurements. Wells says that stress in JC is marked by lack of reduction in unstressed syllables. If we look at the charts in Figures 6.9, 6.10, it turns out that on the contrary, most of the vowels have significant effects of stress.

The upper charts in Figures 6.9 and 6.10 show the effects of stress (more precisely, of destressing) when clitics are excluded, while the lower charts show the effects of stress on vowels in when clitics are included. In some other dialects, including or excluding the clitic words makes the pattern more clear, so both figures are presented. Very little difference is observed between the upper and lower charts for either speaker. So I will discuss just the lower charts, which includes both clitic and non-clitic words, since it contains more data.

I will discuss the effects on long vowels first, and then the effects on short vowels. The most striking fact is that the high and low long vowels have no significant effect of stress for Roasta, and relatively small (though significant) effects for Juba.

We may explain the stress effects on long vowels as an assimilation of gliding vowels to their phonetic off-glides. Thus /ie, uo/ have a strong shift from their stressed nuclei in the direction of their offglides. Note that /ie/ for both speakers glides downward along the edge of vowel space, not inward, towards [ə], consistent with the transcription as /ie/, as opposed to /io/, as might instead have been expected. The other long vowels show insignificant or small effects of stress, perhaps because they have no glides, or phonologically specified glides (that is, not phonetic glides). Long monophthongs /ii, aa, uu/ lack offglides and thus show no significant effect of stress for Roasta, and only small effects for /ii, aa/ for Juba. “True” diphthongs /ai, ou/ have two phonologically specified targets, and thus nucleus targets appear to be attained whether the vowel is stressed or not — if the effect is limited to phonetic diphthongs. This rather subtle distinction between glides and diphthongs has previously been made on phonetic grounds by Lehiste and Peterson (1961). They listed English [ai], [au], [oi] as long, complex vowels with double targets, while the other long vowels were either monophthongs or were classified as “glides” — i.e., as single-target vowels with a phonetic offglide that did not constitute a second phonetic target. This

---

17That is, the Reference American vowels /ay, aw, oy/, or Jamaican /ai, ou, ai/, corresponding to the lexical sets PRICE, MOUTH, and CHOICE.
Figure 6.9: JC Stress Effects (Juba): Stressed and unstressed vowels in non-clitic words only, and in both clitic and non-clitic words. Arrows point from the mean of stressed tokens to the mean of unstressed tokens for each vowel class.

JC (Juba): Effects of Destressing.
Clitics excluded

Dashed arrows are insignificant on 5%, 2-tail t-test

JC (Juba): Effects of Destressing
Clitics included

Dashed arrows are insignificant on 5%, 2-tail t-test

184
Figure 6.10: JC Stress Effects (Roasta): Stressed and unstressed vowels in non-clitic words only, and in both clitic and non-clitic words. Arrows point from the mean of stressed tokens to the mean of unstressed tokens for each vowel class.

JC (Roasta): Effects of Destressing
Clitics excluded

\[ \begin{align*}
\text{F1 (Hz)} \\
\text{F2 (Hz)}
\end{align*} \]

Dashed arrows are insignificant on 5%, 2-tail t-test

JC (Roasta): Effects of Destressing
Clitics included

\[ \begin{align*}
\text{F1 (Hz)} \\
\text{F2 (Hz)}
\end{align*} \]

Dashed arrows are insignificant on 5%, 2-tail t-test

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
phonetic distinction corresponds to a phonological distinction, where the glide is specified underlingly on /ai, ou/, but not on /ie, uo/ (underlingly /e:, o:/, in the lexical sets FACE and GOAT). In the phonological analysis proposed above, the realization of /e:, o:/, as phonetic sequences of high nuclei followed by mid glides, is derived by one of the rules that specifies the surface phonetic facts.

The phonetically derived glides, /ie, uo/, are in a sense less strictly specified than glides which are specified in the underlying phonological form. A reasonable way to look at this data is to consider that the phonetic rules that create the derived glides interact with (phrasal) stress, in such a way that derived nucleus-glide sequences are less phonetically distinct when stress is reduced.

Another way of looking at the same data is to make the analogy to Gallo-Romance breaking, in which vowels diphthongize under stress. That is, rather than viewing the effect of stress reduction as nucleus-glide assimilation, perhaps instead we can view the breaking and raising of the nuclei of underlying /e:, o:/ to /ie, uo/ as a phonetic process that is triggered by the application of stress (see also LYS):

The nucleus of /e:, o:/ is raised and diphthongized when stressed.

If true, this explains why the unstressed nuclei are so much lower than the stressed nuclei: they have not been diphthongized and raised from their underlying positions. This view would also explain why it is that these two vowels, among all the long vowels, have the strongest effect of stress: the process of breaking and raising under stress is restricted to the long, underlyingly mid vowels /e:, o:/, and doesn’t apply to /i:, a:, u:/.

Next let us consider the short vowels. Of these, /i, a, u/ shift a considerable distance towards the center of the vowel space, while /e, o/ shift a relatively small distance towards the same reduction target. Why aren’t the short mid vowels also centralizing? Notice that /e, o/ are already mid vowels to begin with; because they are short and non-peripheral, a process of reduction which shifts vowels to a mid and central position need not move these vowels very far at all. So even if /e, o/ were to centralize, it would have only a slight effect, because they are already mid and relatively central at their stressed, target locations.

---

This makes it reasonable to state that all the short vowels centralize under phrasal stress reduction. We may summarize the effects on the short vowels by the statement,

Centralization due to phrasal stress applies to short vowels.

If we explain the effects of stress on long vowels and diphthongs as suggested above (that is, not by a rule of centralization), then we may summarize the effect of the process of centralization on both short and long vowels by appending “only” to the above statement.

The effects of stress for L.A. Chicano English, and for Chicago White English, are characterized as shifts in the direction of a reduction target. That is, all the arrows, displaying the effect on formant values of stress reduction, point fairly precisely towards a single point, with a small number of exceptions that are explained for other reasons. In Jamaican, the picture is not so clear, until we look at the effects on the short vowels alone, as argued above. The short vowels, /i, e, a, o, u/, in Juba’s and Roasta’s clitics-excluded charts, point towards a single location in formant space (F1 ≈ 420Hz, F2 ≈ 1300Hz for both of them, though this point is relatively higher within the vowel space for Roasta than for Juba). Thus the phonetic process of stress reduction, shifting vowel nuclei in the direction of a single “reduction target”, can be seen to apply in Jamaican Creole as well as in other English dialects, though in JC it appears to be restricted to applying to the short vowels.

6.8 Summary

This chapter has approached the Jamaican Creole vowel system from a number of directions. An analysis of the phonological structure of the JC vowel system was proposed. The overall shape of vowel formant space was compared with that of other dialects. Impressionistic transcriptions of stressed tokens of each of the vowels were listed and briefly discussed. The phonetic effects of phonological vowel length were explored: long vowels have twice the duration of short vowels, and long vowel nuclei are peripheral and shifted into the corners of the acoustic space, relative to their short counterparts. Bootstrap resampling was used to show how precise were the estimates of the mean locations for each vowel. These formant-frequency means were then characterized by an acoustic-phonetic
grammar for vowel nuclei that was considerably simpler than the alternative list of mean formant frequencies. This grammar is similar to the phonetic implementation grammar given for Reference American in Section 3.6, except that it includes a number of dialect-specific phonetic implementation rules. The effects of stress on vowel quality was explored: Centralization due to phrasal stress reduction applies to the short vowels only. The differences between the effects of following /l/ in Jamaican Creole as opposed to other dialects is discussed in Chapter 10.

In Appendix 2, sound-shifts and mergers in Jamaican Creole are discussed. Evidence from other English-based Caribbean creoles is adduced to suggest a particular chain of historical sound-shifts resulting in some of the current phonetic characteristics of Jamaican Creole. Finally, because of the irreversibility of merger, it is argued that one can infer that if a basilectal variety has a merger that an acrolectal variety does not, then the acrolectal variety has existed as long as the basilectal variety.
Chapter 7

Chicago White English

In this chapter we examine the phonetics and phonology of three speakers of the Chicago white vernacular. As in other chapters, a number of issues are examined with respect to this dialect. First, the surface phonological contrasts are laid out; then impressionistic transcriptions of fully stressed tokens are presented. The Northern Cities Chain Shift is described, in which most of the non-rhotic vowels in the system are involved in a wholesale rotation of the vowel space. The overall shape of F1-F2 space is discussed. It is shown that the acoustic realizations of single vowel phonemes can be distributed over the entire vowel space. Next, the effects of stress reduction are examined and discussed with reference to theories of vowel reduction. Finally, the coarticulatory effects of certain following consonants are investigated.

7.1 Characteristics of Speakers and Recordings

The speakers are all young, working-class, white Chicagoans, rather uniform in age and social class. One was interviewed in 1970, and the others almost twenty years later, so
some real-time evidence of sound change in progress is present in this data. In all cases a Nagra tape-recorder was used to make the recording, on 1/4" open-reel audio tape.2

The first speaker, analysed most extensively, will be called Rita. She was interviewed in May, 1990,3 at a Catholic high school just outside the Chicago city limits, where she was a senior at the time. She is white, ethnically Italian, and grew up in a mostly Italian and Irish working-class neighborhood in the mid-west of Chicago proper. She is not as advanced in some of the Chicago sound changes as many speakers,4 but her speech is nonetheless clearly a Chicago vernacular. The entire 25-minute interview was acoustically analysed, including both careful- and casual-style speech.5

The second speaker I will call Jim. He was interviewed in January, 1970, by Benjamin Wald, then a student of Labov's on a field-work trip to Chicago from New York. Jim was 15 years old at the time, and gives the impression, both by his speech style and the stories he tells, of being a street-wise, tough young man. He grew up in Irish and black neighborhoods in the southwest of Chicago proper, and was interviewed in a group hangout with other members of a white street gang. By the admiration he receives from off-mike speakers, it would appear that Jim is the leader of his group. Despite considerable contact with black speakers (he says he was the only white person in his eighth grade class, before his family moved to its current neighborhood), I can identify no distinctively African-American English features in Jim's speech. The non-assimilated nature of his accent fits the pattern of the lack of transmission of dialect features across this ethnic boundary found in Philadelphia (Ash & Myhill 1986). It is clear upon listening that Jim is an impressive vernacular speaker of Chicago White English. Four narratives, were analysed, constituting 11.5 minutes of speech.

The third speaker, here called Judy, was also interviewed by Sharon Ash, in April 1988.

---

2Nagra produced the highest quality portable tape-recorders available before digital audio tape.
3The interview was conducted by Sharon Ash as part of the fieldwork done for the project on Cross-Dialectal Comprehension with William Labov. I would like to express again my thanks to her and to the members of CDC project.
4Speakers from other Northern Cities, such as Detroit and Buffalo, are equally advanced, sometimes more advanced in some of the changes.
5Casual speech is the least-monitored style of speech, in which speakers pay less attention to how they are speaking than to the content of what they are saying. Casual speech is formally defined as speech that is directed to people other than the interviewer and narratives, in which the telling of important events takes precedence over the constrained self-consciousness of the interview situation. Careful speech is all other extemporaneous speech.
She was recorded in a quiet room after a linguistics class at the University of Illinois at Chicago Circle, a commuter school serving urban Chicago. Her parents are Polish and Armenian; she has a passive understanding of Armenian through listening to her father's parents, but claims she cannot speak it, and has no competence in Polish. She was 19 years old, lived with her mother, and worked while attending school. She grew up in Chicago proper, near Wrigley Field to the north of downtown, and exhibits something of a disdain for wealthier suburbanites who "can't scratch two dimes together" because they so commonly use "plastic money" — credit cards. Judy is the most advanced⁶ speaker of the three; some sentences from this interview were used in the experiments of the project on Cross-Dialectal Comprehension as extreme examples of Chicago advanced sound changes. In those experiments it was found that phonetically advanced vowel tokens are confusable with other sound classes by speakers of other dialects. 9.5 minutes of narratives were extracted from the interview and analysed.

A tradeoff was found between the carefulness of the speech and the amount of background noise. Background noise and noises were minimal for Ash's interviews of Rita and Judy, but a larger quantity of speech was needed to get the same amount of narrative speech. On the other hand background noises and overlapping speech in the Jim's interview caused 18% of his tokens to be thrown out, but the speech style was quite un-self-conscious. In no case were measurements taken where they did not clearly reflect a formant in the subject's speech, as shown by listening to the vowel and by visual examination of a spectrogram of it.

7.2 Surface Phonological Inventory of the Vowel System

Here I discuss the set of phonological contrasts in Chicago White English (henceforth CWE). A summary of the next few paragraphs is given in tables of phonemes and the lexical sets that constitute them below. Again, we will use the lexical sets of Wells (1982) as a convenient way of summarizing the phonological differences between this dialect and others.

CWE, like LA Chicano English and most American dialects, does not distinguish the

⁶That is, phonetically advanced along the paths of the sound changes in progress in her dialect.
pairs of Wellsian lexical sets NORTH and FORCE; CLOTH and THOUGHT; BATH and TRAP; PALM and LOT. But CWE is not entirely like LACE in inventory: the merger of /a/(CLOTH, THOUGHT) with /a/ (LOT, PALM) has not overtaken Chicago, though it has reached other cities in the Great Lakes area including Cleveland and Pittsburgh.

Some confusion exists about the “broad A” class of words (Wells’ PALM set). Bloomfield (1934) reported that “the great majority [of educated Chicago speakers] do not distinguish the vowel of hot, sod, bomb from that of father, far, balm” (p. 97), that is, LOT is not distinguished from PALM (so-called “short O”, and “broad A”). However, he claimed that “some speakers, like B[ritish RP speakers], distinguish [these vowels]” (p. 97). Of the Americans that Bloomfield had contact with at the University of Chicago, it is quite likely that many had exposure to RP, which held great prestige in the U.S. at that time, and further that some of these speakers modified their pronunciations of some items in the PALM class in the formal elicitation style which linguists of that era commonly used for data. Thus it cannot be assumed that broad A has ever been distinct in the vernacular English dialect of Chicago or the other Northern Cities.

Further evidence can be found to support this position. An examination of the Northern Cities speakers in Labov, Yaeger, and Steiner (1972, Figures 10-23 and 28) shows that where broad A (PALM, symbolized as /ah/ in those figures, following Trager & Bloch 1941) was classified separately from short O (LOT) (that is, in Figures 11, 14, 16, 18, 28), they always overlap each other, more or less completely, excepting only a single token in a reading-list passage by one Chicago speaker (Fig 28). Thus there appears to be no good evidence that the PALM set is distinguished from the LOT set in Chicago.

The unstressed high-front vowel of the lexical set HAPPY is in this dialect higher and fronter than /i/ as in KIT, and furthermore, upgliding, so that it is phonetically most similar to the vowel /iy/ as in FLEECE. Phonetic similarity and complementary distribution suggests that HAPPY and FLEECE contain the same phoneme. This is different from the situation in British RP, for example, where HAPPY ends in a high-front lax vowel [], making it phonetically most similar to the phoneme in KIT. Thus it is identified phonologically as /i/ in RP. By a similar argument, unstressed /a/ as in

---

7 Including 5 from Detroit, 4 from Buffalo, 1 from Rochester, and 4 from Chicago
8 FLEECE-type words have lexically stressed /iy/, while HAPPY-type words have a lexically unstressed vowel.
LETTER is the unstressed allophone of the /ə/ phoneme in NURSE in this dialect. This differs from partially rhotic dialects such as Jamaican, where NURSE retains some phonetic rhoticity, but where the final vowel in LETTER is not distinguished from the final vowel in COMMA. Thus the following phonemes are contained in words of the indicated lexical sets:

\[
\begin{align*}
/a/ & \quad \text{LOT = PALM} \\
/o/ & \quad \text{THOUGHT = CLOTH} \\
/e/ & \quad \text{TRAP = BATH} \\
/r/ & \quad \text{NORTH = FORCE} \\
/y/ & \quad \text{FLEECE = HAPPY} \\
/ə/ & \quad \text{NURSE = LETTER}
\end{align*}
\]

I assume that the phonological inventory, exclusive of */rei/* and the unstressed vowels, is that represented in Table 7.1.⁹

Table 7.1: Chicago White English Vowel Structure

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>V:</th>
<th>Vr</th>
<th>Vy</th>
<th>Vw</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i</td>
<td>u</td>
<td>iy</td>
<td>uw</td>
<td>ir</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>æ</td>
<td>ey</td>
<td>ow</td>
<td>er</td>
</tr>
<tr>
<td>low</td>
<td>æ</td>
<td>a</td>
<td>o</td>
<td>o</td>
<td>or</td>
</tr>
</tbody>
</table>

As in Reference American, the long vowels are formally represented nuclei with phonologically unspecified glides. The phonetic glides that occur on /iy, ey, ow, uw/ are attributed to rules of phonetic implementation. The -y, -w transcriptions, used in this chapter, represent the more fully specified phonetic forms which are derived from the abstract classes.

⁹The lexical sets corresponding to these phonological classes are:

<table>
<thead>
<tr>
<th>V</th>
<th>V:</th>
<th>Vr</th>
<th>Vw</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT</td>
<td>FOOT</td>
<td>FLEECE</td>
<td>GOOSE</td>
</tr>
<tr>
<td>DRESS</td>
<td>STRUT</td>
<td>FACE</td>
<td>GOAT</td>
</tr>
<tr>
<td>TRAP/BATH</td>
<td>LOT/PALM</td>
<td>Thought/CLOTH</td>
<td>NEAR SQUARE FORCE/NORTH</td>
</tr>
<tr>
<td>marry</td>
<td>START</td>
<td>Unstressed HAPPY</td>
<td>COMMA</td>
</tr>
<tr>
<td></td>
<td>PRICE</td>
<td>MOUTH LETTER</td>
<td></td>
</tr>
</tbody>
</table>
Two further comparisons to Reference American should be pointed out. First, *Mary*, *merry*, and *marry* have merged in Chicago (though Bloomfield’s (1934) discussion suggests otherwise).

Second, despite Bloomfield’s conjectures about broad A in Chicago, it is apparently not distinguished there. The gap opened up by the merger of broad A with short O (PALM and LOT, respectively) in the long, low-front position would therefore seem to be part of the explanation for the lengthening, peripheralizing, and raising of the /æ/ phoneme (TRAP, BATH), which is the first step in the Northern Cities Chain Shift discussed below. The raising of /æ/ is one of the most well-understood American sound changes, both in its phonetic and social conditioning (cf. LYS:Chapter 3, Labov 1990, Callary 1978, Feagin 1991, among others). If this raising is analysed phonologically as first lengthening and then peripheralizing, raising, and breaking, then the first step is the shift of this sound class to the long vowel (V:) subsystem. But this shift itself has as a precondition a gap in the /a:/ slot, which is occupied in Reference American by the broad A class. Thus the merger of broad A and short O (PALM and LOT) is a precondition of the raising of /æ/. Further research on English dialects is necessary to confirm this prediction.

Third, the phonological identification of /a/ with /a/ may be impossible in this dialect, since /a/ (but not /a/) is moving to the low-back corner, as discussed below.

### 7.3 Impressions of Stressed Vowels: The Northern Cities Chain Shift

Two vowel shifts, involving a number of phonemes, are rearranging this vowel system. One is a parallel chain shift, where a set of vowels moves together, in the same direction across vowel space, while the other is the circular drag chain shift called the Northern Cities Chain Shift. In a drag chain shift, one vowel changes to become a different phonetic form, leaving behind a gap in phonetic space, which is filled by another vowel that moves into the gap, leaving behind another gap, which is in turn filled by another vowel, etc., etc.

The movement of the vowels in this chain shift forms a closed loop. A total of 6 vowel phonemes, one after another, have moved so far from their original locations that the new form of one vowel overlaps the old form of another. This is a widespread, ongoing, historical
change in the English vowel system, the Northern Cities Chain Shift (Labov 1991), which is in progress in Detroit, Buffalo, and other cities of the Rust Belt.10

The Northern Cities Chain Shift is as dramatic as the Great Vowel Shift of Early Modern English. It is evident in the following impressionistic transcriptions. The first 10 stressed tokens of each non-rhotic vowel were listened to several times and transcribed with the range of impressionistic realizations presented in Table 7.2.

Table 7.2: Impressionistic Transcriptions of Stressed Vowels (Rita)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Impressionistic Transcriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iy/</td>
<td>[i, i₁, i₁, i]</td>
</tr>
<tr>
<td>/i/</td>
<td>[i, i₁, i₁, i₁]</td>
</tr>
<tr>
<td>/ey/</td>
<td>[i₁, e, e₁, i₁]</td>
</tr>
<tr>
<td>/e/</td>
<td>[e, e₁, a, a₁, a']</td>
</tr>
<tr>
<td>/aʊ/</td>
<td>[a, a₁, a, a₁, a']</td>
</tr>
<tr>
<td>/aʊ/</td>
<td>[a, a₁, a, a', a, a₁, a'']</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>[a, a₁, a, a₁, a']</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>[a, a₁, a, a₁, a']</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>[a, a₁, a, a₁, a']</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>[a, a₁, a, a₁, a']</td>
</tr>
</tbody>
</table>

Comparing these impressionistic values to the Reference American phonetic values, we find a number of differences, which reflect the progress of the Northern Cities shift. /æ/ has raised from low to upper mid, as shown by [i, e, e₁] transcriptions above.11 In the wake of the raising of /æ/, /a/ and /ɔ/ are shifting to the front, as reflected in [a] values for the former, and [a'] values for the latter, except before /l/. Then /ɔ/ appears to be backing and falling, evidenced by some [ɔ] transcriptions. The next vowel in this chain shift is /ɛ/, with some tokens as far backed as [ɛ] among the first ten stressed instances transcribed. Finally, /l/ may be undergoing some backing or centralization, possibly preliminary to a downward shift into the [ɛ] area.

To summarize the Northern Cities Chain Shift as it occurs in Chicago, /æ/ first raises to mid- or high- front; then /a/ and /ɔ/ move frontwards, and /ɔ/ backs and falls towards [ɔ]; last /ɛ/ moves backwards towards [ɛ].12 13 It is also possible that /l/ is starting to

---

10See LYS for the first discussion of this shift that I am aware of.
11Some ingliding does occur with this vowel (4 of the first 30 stressed /æ/ tokens), but none of the first ten sounded ingliding to my ear, perhaps because these were mostly non-final, thus not especially lengthened tokens.
12In some productions /ɛ/ moves downward towards [æ].
13Bloomfield (1934), when he phonemicized the TRAP, BATH classes (here, /æ/) as /ɛ/ , and the STRUT class (here /ɔ/) as /ɔ/, would seem to have been prescient (unless he was influenced by the
back and fall. Thus the pattern is:

\[ (/\text{i}\to)/\text{e}/ \to /\text{a}/ \to /\text{o}/ \to /\text{e}/ \rightarrow [\text{e}^\text{\textsuperscript{\textcircled{\textast}}} /\text{e}]. \]

Advanced tokens of each phoneme in the Northern Cities Chain Shift are easily found in recordings of vernacular speech which, when played in isolation, sound entirely like conservative tokens of the overlapped phoneme. As Veatch (1987), Labov & Veatch (1987), and others have shown in Labov's project on Cross-Dialectal Comprehension, even Chicagoans who speak the advanced vernacular dialect can be completely fooled by such tokens, despite the fact that they are native speakers of this dialect.

This chain shift is an exception to one of the general constraints on chain shifting of Labov, Yaezer, & Steiner (1972). There the observation is made that in chain shifts:
1: peripheral vowels rise; 2: non-peripheral vowels fall; and 3: back vowels move to the front. These principles are said to hold much more strongly of chain shifts than of isolated shifts of single vowels. The Northern Cities Chain Shift, however, includes at least two non-peripheral vowels which move not just down, but also back. The front vowel /\text{e}/ has central realizations, and the central vowel /\text{a}/ has back realizations. Thus this chain shift constitutes a rare exception to principle 3.\(^{14}\)

The Northern Cities Chain Shift is not the end of the changes affecting the Chicago vowel system. All the back vowels except /\text{oy}/ appear to be phonetically fronted.\(^{15}\) Thus we variably find phonetically fronted and even fronted-unrounded tokens of /\text{uw}, \text{u}, \text{ow},\(^{\text{16}}\) \text{\textbeta}/. Clearly not all tokens are shifted, but many are, to a greater or lesser degree.

Yet another vowel may be on the move in Chicago: the vowel /\text{oy}/ was variably transcribed here with a nucleus raised towards \([\text{u}]\). Indeed, only about a third of the vowels are not undergoing change in this dialect, namely /\text{iy}, \text{ay}, \text{aw}, \text{\textsigma}/.

Considering the large implications of these impressionistic transcriptions, let us now go on to the instrumental characterizations of Chicago White English vowels, and see if the impressions are confirmed by the measurements.

\(^{14}\)But cf. recent revisions in Labov, forthcoming.
\(^{15}\)Fronting is relative to the values in Reference American, where /\text{u}, \text{u}/ are realized as \([\text{u}^\text{\textsuperscript{\textcircled{\textast}}} , \text{u}]\).
\(^{\text{16}}\)The vowel /\text{ow}/ is sometimes monophthongal, as is more typical of Minnesota or Wisconsin, but more often has an upglide.
7.4 The Shape of Vowel Space

Three Chicago speakers were analysed, so as to give weight to the claim that the patterns found are characteristic of the dialect, rather than idiosyncrasies. F1-F2 measurements of their vowels are given in Figures 7.1, 7.2, 7.3.

The first speaker analysed, Rita, is the subject of one of the larger acoustic analyses of a speaker’s vowel system, with 4821 vowels examined, and 4479 measured. 17542 were not measured either because of extraneous noises (e.g., overlapping speech) that degraded the signal, or due to deletion or devoicing of the vowel. F1-F2 measurements of all tokens for the three speakers are displayed in Figures 7.1, 7.2, 7.3.

The F1-F2 vowel spaces are somewhat different from one another; the most obvious visual difference is the overall density of points on each graph, which is due simply to the different numbers of tokens measured. The three charts share fundamental properties. The distributions are roughly triangular, with somewhat of an asymmetry between the front and the back: the high-back corner is relatively sparsely filled. This reflects the frequencies of the various vowels: high-back vowels are less frequent than other vowels for all the Chicago speakers. These frequencies are shown in Table 7.3, where the counts are ordered from most to least frequent for each speaker.

Vowels between horizontal lines vary in relative frequency across speakers; sets separated by horizontal lines do not vary in relative frequency in this data. In particular, notice that the high-back vowels /uw, u/ are among the least frequent vowels for all speakers. This accounts for the lower density of tokens in the high-back corners of these vowel spaces.

Similar factors may be responsible for the front-back asymmetry found also in Chicano, Alabama, and Jamaican Creole. The main difference between these charts and similar charts made for other dialects (pages 166, 167, 233, 254.) is that these are more perhaps more evenly distributed over the entire space. The Chicano speaker’s tokens are more concentrated in the upper half of the space, while the Alabama speaker very sparsely fills both the high and low regions, and the Jamaicans have a mode in the low corner and (for

---

17 In an important study of the effects of situational style-shifting on a speaker’s vowel system, Hindle (1980) measured even more tokens (> 10,000) in the speech of a single Philadelphia speaker, recorded during work, at dinner with her family, and during a bridge game with friends. The situation-related vowel shifts are closely related to the historical changes ongoing in Philadelphian sound system, in that sounds in more vernacular, informal-style speech is more historically advanced.
Figure 7.1: 4470 F1/F2 measurements. Interviewed 1990.

Chicago Vowel Space: Rita, 17.
Figure 7.2: 1614 F1/F2 measurements. Interviewed 1988.

Chicago Vowel Space: Judy, 19
Figure 7.3: 2328 F1/F2 measurements. Interviewed 1970.

Chicago Vowel Space: Jim, 15.
Table 7.3: Phoneme counts ordered by frequency for each speaker. Horizontal lines distinguish frequent, common, and infrequent vowels.

<table>
<thead>
<tr>
<th></th>
<th>Rita</th>
<th>Jim</th>
<th>Judy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ɪ</td>
<td>564</td>
<td>ə</td>
<td>266</td>
</tr>
<tr>
<td>ay</td>
<td>555</td>
<td>ay</td>
<td>325</td>
</tr>
<tr>
<td>ə</td>
<td>543</td>
<td>ɪ</td>
<td>259</td>
</tr>
<tr>
<td>ʌ</td>
<td>427</td>
<td>ow</td>
<td>170</td>
</tr>
<tr>
<td>iy</td>
<td>491</td>
<td>iy</td>
<td>167</td>
</tr>
<tr>
<td>ɛ</td>
<td>352</td>
<td>əɛ</td>
<td>149</td>
</tr>
<tr>
<td>ow</td>
<td>341</td>
<td>ʌ</td>
<td>140</td>
</tr>
<tr>
<td>æ</td>
<td>278</td>
<td>ɛ</td>
<td>108</td>
</tr>
<tr>
<td>ø</td>
<td>220</td>
<td>ey</td>
<td>108</td>
</tr>
<tr>
<td>a</td>
<td>178</td>
<td>a</td>
<td>101</td>
</tr>
<tr>
<td>ey</td>
<td>177</td>
<td>ø</td>
<td>94</td>
</tr>
<tr>
<td>uw</td>
<td>169</td>
<td>aw</td>
<td>62</td>
</tr>
<tr>
<td>ɔ</td>
<td>75</td>
<td>ɔ</td>
<td>51</td>
</tr>
<tr>
<td>aw</td>
<td>50</td>
<td>u</td>
<td>45</td>
</tr>
<tr>
<td>u</td>
<td>40</td>
<td>uw</td>
<td>37</td>
</tr>
<tr>
<td>oy</td>
<td>5</td>
<td>oy</td>
<td>1</td>
</tr>
</tbody>
</table>

Juba especially) a large gap in the high-central region. These interesting and suggestive patterns of the overall envelope of variation for other dialects can best be interpreted with reference to the phonological system of the dialect.

### 7.5 Overlap of Phonemes

An important lesson here is that the phonetic categories do not simply jump off the page when examining measurements of individual tokens. The data found here show tremendous overlap between different phonemes. In the most extreme example I found, the distribution of the /ɛ/ phoneme for Rita is nearly coextensive with the entire vowel space, as shown in Figure 7.4.

Lisker (1949) found some irreducible overlap between F1 and F2 measurements for /ɛ/ and /æ/ in a particular context (p-p). (A replication of his thesis was conducted, and is described on pages 118ff.) Here we find that /ɛ/ overlaps not just /æ/ but all the vowels in the entire space. The dispersion of /ɛ/ is the most extreme case in Rita's data. (It is also, perhaps not coincidentally, the most recent vowel to join the chain shift.) However,
Figure 7.4: Rita’s Vowels; Rita’s /E/

all vowels

/F2 (Hz)/

/E/ only

/F2 (Hz)/

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
many vowels in Rita's system spread over nearly as much of the space, and few cover less that half of the entire area. For similar results in another dialect, see the charts of Juba's (Jamaican Creole) long and short vowels, on page 172. The striking amount of dispersion shown in Figure 7.4 is not spurious or mistaken. Because each of the tokens was examined individually, and formant tracks were corrected,\textsuperscript{18} the number of erroneous measurements are few: the dispersion of /e/ tokens in Figure 7.4 cannot be attributed to gross errors of measurement. Thus we may conclude that when a single speaker's natural conversational speech is analysed, including both stressed and unstressed tokens and all the coarticulation and reduction to be found in such contexts, the amount of overlapping between vowels is extreme. The problem Lisker (1949) points out becomes much worse under these conditions.

The fact that the nucleus formant frequencies are a very reduced representation of the vowel tokens studied leads us to speculate that overlapping phonemes may be distinguished by other factors which are not present in these charts, such as glides in one direction or another, differences in vowel duration, F3 frequency differences, even formant bandwidth or amplitude differences. Information about the the stress level and phonetic context of the sounds measured, plus dialect-specific knowledge about the contexts and rules for natural-speech reduction, may be sufficient to disambiguate most tokens on the basis of phonetic and phonetically derived information alone. Overlap is not necessarily an insoluble problem for classification. It is quite possible, even likely, however, that a residue of unclassifiable tokens will remain, even after applying all the phonetic knowledge possible. Some sounds are simply ambiguous, and listeners must bring higher-level linguistic knowledge to bear on the decoding problem, or fail in their efforts.\textsuperscript{19}

In fact, since our primary concern here is with phoneme-internal variation, this is actually, for present purposes, a good situation. The more variation to be found in the pronunciations of a single phoneme, the better we will be able to substantiate the sub-phonemic alternations or the allophonic effects which are the focus of this work. This is an important reason that natural speech is superior to laboratory speech for studying effects of coarticulation and reduction. More processes apply, more frequently, in natural speech.

\textsuperscript{18}See page 146ff for discussion of the post hoc examination of outliers.\textsuperscript{19}Failure to decode the sounds of speech does occur; language does not always function successfully, as Labov's project on Cross-Dialectal Comprehension shows in detail.
Contextual variation is explored in Chapter 10. This source of variation as well as time-varying spectral information are essential to recognizing the phonological representations corresponding to natural speech. The next section of this chapter discusses the mean locations of vowels in F1-F2 space, the locations from which vowels deviate when affected by particular influences of stress and phonetic context. The subsequent section investigates the influence of stress, and the last section examines the effects of the consonants in the adjacent phonological context.

7.6 Bootstrapped Mean Distributions and Sound Change

The bootstrap technique provides an estimate of how precisely located are the means of the distributions of F1-F2 measurements for each vowel. The bootstrap is discussed in detail in the Methods chapter. The clouds in Figures 7.5, 7.6, 7.7 represent the intrinsic scatter in the estimate of the mean of each vowel’s distribution. The true mean is some undetermined point within the cloud of points; if two clouds containing a couple hundred points each do not overlap each other, then it is quite unlikely that they could have the same mean.

Figures such as these can most easily be interpreted by observing the relative positions of vowels in F1-F2 space, and comparing these relations to other known relations. Absolute frequencies are quite difficult to interpret (though not impossible, given information about the speaker’s vocal tract dimensions, and given considerable experience in relating measurements to sound qualities). This discussion will focus on the clearly observable relative positions of vowel phonemes in this space.

7.6.1 Evidence for the Northern Cities Chain Shift

The relative positions of the vowels in these figures show the changes ongoing in the Northern Cities Chain Shift. The raising of /æ/ and the lowering and backing of /ɛ/ are evident in Judy’s and Rita’s charts, since their relations are reversed as compared with the expected locations for Reference American (/ɛ/ is higher than /æ/ in RA), or for example, the relation seen between these vowels in Los Angeles Chicano speech (Figure 9.2, page 256). While /ɛ/ is evidently backed for Jim (it is located as far back as the central
Figure 7.5: Bootstrapped mean distributions.

CWE (Rita): 200 Re-estimated means for each vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CWE (Jim): 200 Re-estimated means for each vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 7.7: Bootstrapped mean distributions.

CWE (Judy): 200 Re-estimated means for each vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
low vowels and /æ/, it is not apparently lowered in his phonetic vowel system, though /æ/ has raised to the same height as /ɛ/. This is consistent with the view that the Northern Cities Chain Shift has progressed since 1970, when Jim was interviewed. /æ/-raising may be the first step in the chain shift,\(^{20}\) while the lowering and backing of /ɛ/ to [æ] and [ʌ] is among the most recent changes. So if Jim’s /æ/ is raised, but his /ɛ/ isn’t lowered as much as /ɛ/ is lowered in Judy’s and Rita’s vowel systems, this is because Jim represents an intermediate stage of the chain shift.

The fronting of /æ/ is perhaps evident in the fact that for each of the three speakers, /æ/ lies to the front of /aw/ and is distinct from it. /æ/ is the most low-front vowel in the system for all of them. That this is a drag chain shift is evident from the fact that /æ/ is fronted and raised a great distance away from /æ/, leaving a large phonetic gap for /æ/ to move into.

The fronting of /ɔ/ proceeds gradually, from Jim, for whom /ɔ/ is well to the back of /ar, aw, a, ay/, to Rita, for whom /ɔ/ is quite close to /ar, aw/, to Judy, for whom /ɔ/ has become the lowest vowel in the system, and is almost indistinguishable (in terms of average nuclei in F1-F2 space) from /aw/.

The approximation of /ɔ/ with the /a-/ vowels, /a, aw, ar/, in Judy’s speech raises the question of merger: Will Chicago undergo the Low-Back Merger? It has not apparently done so yet, but the difference between /a/ and /ɔ/ is quite small for Judy.

The lowering of /ʌ/ (which precedes /ɛ/-lowering, discussed above) is evident by its relationship with /ow/, which is much higher than /ʌ/ for Judy and Jim, though not for Rita.

Thus each of the steps of the Northern Cities Chain Shift, as well as the fronting of the high back vowels, /uw, u/, are evident in Figures 7.5-7.7.

The fronting of the high back vowels, /uw, u/ is not evident for Jim, where these vowels are as far back as /ow, or/. But in both Judy and Rita’s charts, /uw, u/ are well to the front of /ow, or/, again reflecting the progress of sound change in the intervening 20 years.

\(^{20}\)However, according to the discussion on page 101, /æ/-raising is the second step, which follows the loss of the Reference American /a:/ phoneme, broad A, by merger with /ɑ/. It was shown above that /a:/ is not a distinct class in Chicago.

208

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
7.6.2 Short Vowel Lowering and Centralization

For all three speakers, /1, ə, ʊ/ lie in a row across the non-peripheral upper edge of the vowel space, each different from the other. This is evidence that the proper form of phonological schwa is /ɪ/, a short, central, high vowel.

For all three speakers, the non-low short vowels, /1, ɛ, ʊ, ʌ/, are centralized and lowered, relative to their long counterparts /iy, ey, uw, ow/; the back vowels are not centralized (fronted) as much as the front vowels are centralized (backed). Note that for Rita and Judy, who have a fronted /uw/ and /ʊ/, this effect takes the form of lowering, since these nucleus are high-central. The only exception to this rule is Rita's /ʌ/ which is not significantly lower than her /ow/. Apparently Rita is not participating in the lowering of /ʌ/ in the Northern Cities Chain Shift.

This data therefore provide fairly strong support for the the generalization that the non-low short vowels are centralized and lowered relative to the corresponding long nuclei. This generalization is quite similar to the relationship between long and short vowels in Jamaican Creole, formalized in rules 4, 7, pages 178, 179. Rule (7), which states that the short vowels are lowered relative to the long vowels, cannot be a universal fact, since Rita's /ʌ/ is not lower than the corresponding long vowel, /ow/. In fact, the nuclei of long vowels may be identical with those of their short counterparts, as are, for example, /ey/ and /ɛ/ in the Alabama speaker's vowel system (see Figure 8.2, page 236).

Phonetic Grammar Specifies Different Nuclei for /a/- Vowels.

Consider the relations of the nuclei of the low diphthongs /aɪ, aʊ/ to the low vowel /a/, in the above figures. /aɪ/ is raised and to the front of /a/, while /aʊ/ is raised and to the back, for all these speakers. How shall we interpret these relations? It appears that there are three distinct phonetic specifications for the three vowel nuclei. This is unlike the pattern of Reference American, where the nuclei of all three vowels are phonologically identical, that is, all have the same low, back, unrounded nucleus, and where no phonetic rules are postulated to distinguish between them.

An attempt may be made to attribute these differences to natural phonetic effects of the offglides, thereby allowing the same unitary treatment of the three nuclei as in
Reference American. For example, one might suppose that there is a process of nucleus-glide coarticulation or assimilation, by which the opening gesture associated with the nucleus is coarticulated with the closing gesture associated with the offglide. This would explain why /ay/ is raised to the front and /aw/ is raised to the back: /-y/ is high and front and /-w/ is high and back, relative to the /a/ nucleus.

However, if this supposition were true, it would seem to weaken explanations of vocalic sound changes which rely on an explanatory process of nucleus-glide differentiation. If the process of nucleus-glide differentiation “explains” some kinds of variation, and nucleus-glide assimilation “explains” others, it would seem that there remains only description, and no explanation at all. On the other hand, if the two processes were restricted to complementary environments, we could maintain the explanatory value of both, within their restricted contexts of occurrence. For example, if assimilation were typical of unstressed environments, while differentiation was typical of stressed environments, then appealing to both processes in different cases remains possible and has explanatory value.

Unfortunately, the effects of stress in the current case do not bear out this explanation. If the three nuclei overlapped when stressed, and separated only when unstressed, then the explanation would be supported. Instead, /ay, a/ rise and come together when unstressed, while /aw/ fronts insignificantly (cf. Figure 7.8). The facts are opposite to the prediction, so the explanation of nucleus-glide assimilation must be abandoned, until some other explanation can be found for these significant phonetic differences. I therefore conclude that the three vowels /ay, a, aw/ indeed have three different phonetic nuclei specified in the phonetic grammar.

Further Observations

The /r/-gliding diphthongs fall into two groups according to how they pattern in Rita's speech, shown in Figure 7.5. For /iy, ey, a/, the corresponding Vr vowels /ir, er, æ/ are located about 300Hz to the back (lower F2 frequency), and 50 to 100Hz lower (higher F1 frequency). For /a, ow/, on the other hand, the corresponding Vr vowels /ar, or/ are located higher and to the back. /ar/ is significantly raised and backed relative to /a/, while the mean measured nuclei for /or/ are barely different from those for /ow/ (the

---

21Chapter 3 argues that postvocalic /r/ is a glide, and forms diphthongs with preceding nuclei.
distributions of re-estimated means overlap, but just slightly, so the difference is probably genuine).

Jim’s /a, ar/ form a single cloud, and thus share an identical nucleus; /ay/ is raised, while /aw/ is backed from /a/. (His /ir/ was not shown in Figure 7.6 in order to keep the chart readable; it lies between and overlapping with /ey, ɪ, ɜ/.)

In all three charts, /ar, or/ are back-raised from /a, ow/ (/ar/ insignificantly so for Judy and Jim). /ir, er/ are insignificantly different for Judy but well separated for Rita; this may reflect insufficient data to show Judy’s distinction clearly.

/oy/ is not shown in Rita’s chart because it occurs only 5 times, not enough to give a precisely located distribution of bootstrapped means. If plotted, this distribution extends from /a/ to the high-back corner, overlapping the realizations of /ow, or, ɪ/. For Judy, the estimated mean for /oy/ may be anywhere from the low-back corner to mid-back. Insufficient numbers of tokens make estimating a mean for this vowel difficult.

An illustration of a limitation of the bootstrap technique (and other statistical methods) was made by applying it, facetiously, to Jim’s single token of /oy/. No matter how many times a sample of one is resampled, estimates of the mean taken from the resampled data are invariant. What appears to be a single “+” is actually two hundred re-estimated means laid on top of one another, taken from this sample of one data point, an obvious misuse of the bootstrap technique. (Similarly, the standard deviation of a sample of one is always zero.) Statistics require reasonable amounts of data.

The distribution of reestimated means for syllabic /ɪ/ is also not shown. For Rita, it overlaps /ow, or/ and ranges up toward the high-back corner. It does not overlap /u, ə, uw/ and other vowels located to its front.

7.7 Stress Reduction

Next we examine the effect of impressionistically coded phrasal stress reduction on instrumental measurements of formant frequencies. As discussed in the Methods chapter, the “stress” I am referring to is not lexical stress as may be listed in a dictionary or derived by a system of rules, but impressionistically classified phrasal stress. Phrasal stress here
means an impressionistic classification of each syllable in the context of the entire utterance in which it was spoken. Vowel tokens in clitic words\textsuperscript{22} are distinguished, and vowel classes with one or no tokens of either stressed or unstressed classes are excluded.

7.7.1 Kinds of Reduction

Theoretical background for this study of vowel reduction is discussed in Chapter 4. There are many possible effects of stress reduction on vowel nuclei:

1) No effect
2) Raising
3) Lowering
4) Fronting
5) Backing
6) Shift towards a mid-central reduction target
7) Shift towards a high-central reduction target
8) Shift towards any given reduction target

To determine which of these are closer to the truth, consider the charts of stressed and unstressed means in Figures 7.8, 7.9, 7.10. In these figures, the tails of the arrows are the means of the stressed tokens; the heads are the means of the unstressed tokens.

The two charts in each figure are only slightly different; their purpose is to test the hypothesis that the inclusion of vowels in clitic words might have a distorting effect on the vowel-reduction pattern. The patterns seem more clear for Jim and Judy: vowels shift fairly consistently in the direction of a single central, upper-mid location, which for Jim is in the area of the realization of unstressed /a/. The picture is most precise for Judy, but the pattern fairly consistently applies to Jim’s vowels also. Rita’s pattern is more clear with clitic words excluded; all the significant effects are in the direction of high-central /a/, excepting the effect of reduction on /er/. Rita’s pattern, when clitics are included, becomes more murky, since /æ, er/ undergo backing rather than shift in the direction of a common reduction target. The greater precision of the pattern found for Judy and Jim may be due to the fact that the speech analysed for these two speakers was casual-style

\textsuperscript{22}“Clitics” are defined in the Methods chapter.
Figure 7.8: Arrows point from mean of stressed tokens to mean of unstressed tokens. ' means that only stressed tokens of that category occur.

CWE (Rita): Effects of Destressing. Clitics excluded

Dashed arrows are insignificant on 5%, 2-tail t-test

CWE (Rita): Effects of Destressing. Clitics included

Dashed arrows are insignificant on 5%, 2-tail t-test

213

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 7.9: Arrows point from mean of stressed tokens to mean of unstressed tokens. ' means that only stressed tokens of that category occur.

CWE (Jim): Effects of Destressing. Clitics excluded

Dashed arrows are insignificant on 5%, 2-tail t-test.

CWE (Jim): Effects of Destressing. Clitics included

Dashed arrows are insignificant on 5%, 2-tail t-test.

214
Figure 7.10: Arrows point from mean of stressed tokens to mean of unstressed tokens. ' means that only stressed tokens of that category occur.

CWE (Judy): Effects of Destressing.
Clitics excluded

F2 (Hz)
Dashed arrows are insignificant on 5%, 2-tail t-test

CWE (Judy): Effects of Destressing
Clitics included

F2 (Hz)
Dashed arrows are insignificant on 5%, 2-tail t-test

215

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
speech. That is, the data for these two speakers was taken from narratives of personal experience, in which self-monitoring is minimized. The data analysed for Rita, on the other hand, constituted the entire 25-minute interview, and much of it was careful-style speech. This may be one reason that the patterns of stress reduction for Jim and Judy are quite clear, while Rita’s pattern has more significant shifts that don’t fit the pattern.

Clearly there are strong effects in this data, so (1) above is ruled out. Most low vowels rise, but most mid and high vowels do not, so (2) can also be ruled out. (3) is a quite unlikely model for vowel reduction, since very few vowels undergo lowering. (4) applies to the phonologically back vowels for Judy and Jim, but its opposite (5) applies to the phonologically non-low front vowels for Jim and to all front vowels for Judy. (6) is incorrect, since the high vowels do not lower to phonetically mid height. (7) seems to be an accurate description of the overall pattern for both Judy and Jim, and also for Rita’s non-clitic words. Most of the arrows for both Jim and Judy display shifts quite precisely in the direction of a single location. The arrows here suggest that there is a “reduction target” toward which vowels shift in this dialect. The phonetic quality of this target is high and central, a quality that may be written as [i]. Is it a coincidence that the apparent reduction target is where it is, namely, in high-central position? One might generalize from this pattern and propose that the universal reduction target is high and central. To be precise, vowel reduction might be characterized as a universal phonetic process of gradient shifts in the phonetic quality of vowel nuclei in the direction of a high-central reduction target. In short, the process might be properly characterized universally as (7), the general tendency of raising and centralizing towards a high-central target. However in the Jamaican data examined in the last chapter, it was shown that a similar reduction target for the short vowels explains the Jamaican vowel shifts that are due to stress reduction. This reduction target was mid and central, not high and central, as found in the data for these speakers. Thus if the process of vowel reduction as shift towards a reduction target is universal, the exact phonetic target appears not to be. So far we have evidence from two dialects that patterns (6) and (7) above represent the possibilities for vowel reduction. The last possibility, (8), would seem to be too general and unrestricted a statement, since not just any point in vowel space will do. In the next two chapters we will see if more possibilities need to be added, and begin to further explore the space of possible reduction targets.
7.7.2 Do Stressed Vowels Attain a Target?

One might suppose that stressed vowels are more likely to attain the phonetic target of the vowel phoneme than unstressed vowels. Unstressed vowels should be more reduced and more coarticulated with adjacent sounds, which amounts to being more spread out over vowel space than their stressed counterparts. This section tests this supposition, the "Stressed-Target" hypothesis, in a simple way, with Rita's data.

If stressed vowels reach a target more than unstressed vowels, then their nuclei should be more tightly distributed across vowel space, while unstressed vowels should be more spread out, since they are more likely to be modified by adjacent sounds, etc. In order to compare the scatter in these distributions, then, Table 7.4 shows the standard deviations of F1 and F2 for stressed and unstressed tokens of each vowel.

Table 7.4: Standard Deviations in Hz of F1, F2 for stressed and unstressed vowels (Rita)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stressed</td>
<td>Unstressed</td>
</tr>
<tr>
<td>A</td>
<td>148</td>
<td>147</td>
</tr>
<tr>
<td>ow</td>
<td>108</td>
<td>130</td>
</tr>
<tr>
<td>iy</td>
<td>84</td>
<td>98</td>
</tr>
<tr>
<td>æ</td>
<td>130</td>
<td>129</td>
</tr>
<tr>
<td>ay</td>
<td>127</td>
<td>112</td>
</tr>
<tr>
<td>e</td>
<td>146</td>
<td>151</td>
</tr>
<tr>
<td>æ</td>
<td>84</td>
<td>111</td>
</tr>
<tr>
<td>I</td>
<td>115</td>
<td>128</td>
</tr>
<tr>
<td>a</td>
<td>143</td>
<td>133</td>
</tr>
<tr>
<td>uw</td>
<td>124</td>
<td>128</td>
</tr>
<tr>
<td>c</td>
<td>126</td>
<td>163</td>
</tr>
<tr>
<td>ey</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>u</td>
<td>79</td>
<td>108</td>
</tr>
<tr>
<td>aw</td>
<td>125</td>
<td>72</td>
</tr>
<tr>
<td>l</td>
<td>148</td>
<td>147</td>
</tr>
<tr>
<td>oy</td>
<td>108</td>
<td>130</td>
</tr>
</tbody>
</table>

If the stressed vowels are less widely scattered than the unstressed vowels, then the stressed standard deviations should be smaller than the unstressed standard deviations. The claim is that a target is more likely to be achieved in a stressed vowel than in an unstressed vowel. These data give only mild support to this theory. 11 of 17 vowels
(about 2/3) have a smaller F1 standard deviation when stressed than when unstressed; a
different subset of 11 of the 17 have a smaller F2 standard deviation when stressed than
when unstressed. So about two-thirds of the vowels are more tightly distributed in each
dimension when stressed than when unstressed.

However, since about a third of the vowels fail to observe the predicted pattern, the
stressed-target hypothesis is at best a fairly weak generalization.

7.8 Summary

In this chapter I discussed the surface phonological structure of Chicago White English and
presented a number of results of a large-scale acoustic analysis of vowels in the speech of
three speakers of the Chicago white vernacular. All vowels occurring in lengthy, continuous
sections of tape-recorded sociolinguistic interviews were examined, measuring all occurring
acoustic vowels. Vowel quality as measured by formant frequencies at vowel nuclei was
shown to reflect the progress of the Northern Cities Chain Shift, which is shifting the low
vowels and the short mid vowels in this dialect in a circle. The acoustic effects of phrasal
stress were also shown to follow a clear and rather precise pattern, of phonetic shift of
the nucleus of vowels in the direction of a high-central reduction target, which appears to
be different from the target that Jamaican short vowels shift towards. These patterns are
part of the system of phonetic implementation by which abstract phonological categories,
quite closely related to those of Reference American, are realized by patterns of phonetic
form which are characteristic of the vernacular dialect spoken in Chicago.
Chapter 8

Alabama English

This chapter presents analyses of the surface phonology and phonetics of the vowel system of a variety of Southern States English found in Anniston, Alabama. As in other chapters, I discuss the vowel system's surface inventory and structure, some impressionistic transcriptions, the overall shape of vowel space, the sound changes that relate this dialect to others, and the effects of phrasal stress on vowel. This chapter is an extension of Feagin's research on this variety, which concerns including the linguistic and social patterning of tense, mood, aspect, agreement, and negation forms (1979), of non-prevocalic /r/ and other vowel-related changes (1990), of the up-gliding and in-gliding of /æ/ (in press), and other reports, using the speech of this speaker and others interviewed by Feagin between 1968 and 1973. Unlike her work, however, this study is in large part non-developmental. Social and historical differences are discussed only incidentally here. The phonological analysis is based largely on previous published work on Alabama and Southern English, while the phonetic analysis considers measurements of 1637 vowels taken from the conversational speech of a single speaker from Anniston.

8.1 Characteristics of the Speaker

The speaker studied in this dialect was an older, working-class white male with a rural background and little formal education who lived all his life in the Anniston area. He was interviewed with his wife in his home near Anniston in 1972, at the age of 81, by Crawford Feagin, who kindly gave me access to her recordings. He appears in her (1979)
book as James Hays, interview #29. He is a deliberate and entertaining speaker, who tells stories about hunting squirrels and wild turkey, about strikes and the dangers of work in the foundry, and about why he “didn’t care nothin’ about” going out to (square-) dances. His life was typical of the 20th-century, industrial Southern U.S.

Feagin (1990:131) shows that older working-class speakers in this sample are mostly r-ful, as are most young speakers generally. She also shows (in press) that older rural males have relatively less of the Southern drawl (evidenced, for example, in up-gliding and in-gliding of /æ/) than women of their class. This simplifies the analysis below, since we may assume the presence of underlying /r/ as well as surface [œ, œ], and since the considerable complexities of Southern gliding are minimized.1 It must be pointed out that this speaker represents past forms of this dialect, not necessarily current forms, since the dialect of his community has changed considerably since he grew up. It is also clear from Feagin’s work that the amount of sociophonetic variation in the Anniston area was and remains considerable. James H. would be over 100 years old if he were alive today. This does not, of course, affect the theoretical interest of the phonetic effects found in his speech, which we may presume are characteristic of some dialect, namely that of speakers of his age and social background.

8.2 The Surface Phonology of Vowels

This section proposes, discusses, and modifies a structural analysis of the Alabama vowel system. It is not concerned with morphophonological alternations or lexical phonology, but with the structure of the sound system which is the output of the lexical phonology

1Feagin (1991) summarizes the general social picture regarding the disappearance of Southern r-lessness and gliding:

Anecdotal evidence supports this hypothesis. A young upper-class boy from Anniston — the son of a banker — who was attending Amherst College in Massachusetts told me recently that people often are surprised that he is from the South. That is probably because of the stereotype of Southern speech which is that Southern States English is R-less and full of diphthongs on vowels which in Northern States English have none. Actually, the stereotype is not altogether incorrect, for a few older working class people, but it is out of date. The majority of young Southerners — of whatever social class — have R’s now. What the stereotype is depicting for the glides turns out to be working class speech, at least for younger people — or the speech of older women, of whatever class.
and with the ways in which those structural representations are implemented in actual speech. Thus the object of study is surface phonology, or post-lexical phonology in the sense of Kiparsky (1982), as well as phonetic implementation.


Foley (1972) impressionistically studied the phonetics of vowels in Tuscaloosa County, Alabama, using dialect atlas interview techniques. Tuscaloosa is at about the same latitude as Anniston and Birmingham, but lies on the western side of the state. Anniston is in the east of the state, somewhat north of the South/South Midland isogloss of Wood (1961:12) and Foscue (1971:41) that runs east-west through the middle of the state, south of Birmingham. Tuscaloosa is about the same distance north of the Southern/South Midland isogloss as Anniston is, so it could be expected to share most phonological features with Anniston.

The following tabular structure summarizes the phonological vowel classes which I have derived from Foley’s and other work on Southern States English.²³

As in the discussion in Phonological Preliminaries, I exclude prevocalic glides as in cute, quit, as well as vowel sequences that may be analysed as bisyllabic, such as ruin, fluid, Ayre, mayor, lion, Brian. The symbols used in this table, as in other chapters, are not especially phonetic and are not to be taken as IPA transcriptions of these classes. In particular, the V (phonologically short) vowels /1, ɛ, ʊ, etc./ typically have inglides in

²Based on distributions of lexical items rather than phonological features.
³To ease the usually difficult task of understanding what sounds are being referred to, I present here the lexical sets of Wells (1982) that correspond to the sound classes in the given structure. (For further explanation, see the section on English Lexical Sets in Phonological Preliminaries.)
Table 8.1: Alabama Vowel Structure

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>Vy</th>
<th>Vw</th>
<th>Vr</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>ı</td>
<td>u</td>
<td>iy</td>
<td>ı</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>a</td>
<td>ey</td>
<td>ow</td>
</tr>
<tr>
<td>low</td>
<td>æ</td>
<td>a</td>
<td>ay</td>
<td>aw</td>
</tr>
</tbody>
</table>

stressed monosyllables; /ay/ is commonly a monophthong [a:], and the /-y, -w/ symbols in the Vy, Vw classes do not represent phonetic glides whose endpoints are higher than IPA [i, u], respectively, etc. The symbols used here are intended to make reference to locations within the structure simple and unambiguous.

The following discussion describes the relationships of the structure in Table 8.1 to that of Reference American, justifying the different choices made in the structural analysis of this dialect, as compared with that reference dialect. I do not mean to imply that Southern dialects are to be analysed as derived from Northern varieties. In fact, there are more phonological distinctions made in some cases in the South than in the North, so a historical derivation might well go in the opposite direction. However, rather than argue for each detail of an analysis of this dialect from first principles, I have found it simpler to begin with an established system, and modify it as necessary in order to accommodate the partly different and partly similar facts of this dialect. There are a number of notable differences between this (surface phonological) structure and that of Reference American (RA), presented above in Phonological Preliminaries. To summarize: First, among the Reference American long vowels (the V: subsystem), the PALM class (RA /a:/) appears to be merged in this dialect with the LOT class (RA /a/); the THOUGHT class (RA /ɔ:/) is a back-raising glide. The Reference American V: subsystem of /i:, e:, a:, ɔ:, o:, u:/ has no low counterparts in this system, since /a:, ɔ:/ are lost by merger and gliding. The result is a complementary distribution between long vowels and Vy, Vw vowels: none of

---

*It has been claimed (for example, by C.-J. Bailey 1985:205) that as dialects of English, Southern and Northern (and by extension, Reference American) share an underlying vowel system. The systems under discussion here are surface-phonological, or post-lexical in the theory of Lexical Phonology, and they are not the same across these dialects. Surface phonological structures are the output of the lexical phonology, and their details of implementation are assumed not to be subject to the direct influence of morphological or grammatical features in specifying their phonetic realization. The topic here is not lexical phonology, but post-lexical phonology."
the former are low, while all of the latter are low. I therefore shift the mid and high long vowels /iː, eː, oː, uː/ into Vy and Vw subsystems, thereby eliminating the entire class of surface-phonological long monophthongs and increasing the number of phonological glides in Southern speech relative to (Northern) Reference American.

Second, the Vy vowels are found to be slightly different positions within the Vy subsystem. Third, the Vr subsystem is both phonologically and phonetically different. By way of justifying the choices made in proposing this particular structure for Alabama English, I will discuss each of these differences below, first the long vowels and the Vy, Vw gliding vowels, and then the vowels before /r/.

The "broad A" class, RA /aː/, as in PALM, has apparently merged with the low-back checked vowel /a/ as in LOT. Thus, the PALM-class words calm, father words "are assigned to the checked vowel /a/, though the phones are prolonged and frequently backed, as [aː]" (Foley 1972:31).

Foley seems unwilling to assign a set of vowels to a distinct phonological class when it is distinguished primarily by duration rather than by quality. Since a difference of length can be sufficient for a phonemic distinction, this doesn't seem necessary: PALM could be distinct in this dialect. However, McDavid (1940) in a study of the low-back vowels of the Carolina Piedmont, says that the broad A class is not distinct in that closely related dialect. Further research seems called for in light of the length difference found by Foley; for now, I will assume his analysis is correct in assigning broad A to the /a/ class. If the classes are distinguished, then broad A would fill one of the empty low-vowel slots in the V: subsystem.

The THOUGHT vowel class, which is represented in RA as a long low-back vowel, /ɔː/, is a back-raising diphthong in this dialect, most commonly [ɔː, øː], according to Foley as well as in my data, so I have shifted it from its location in the RA V: subsystem to the Vw subsystem, a step which is structurally sensible given that the nucleus of RA /æw/ has fronted to [a, æ] for Foley, and even higher and fronter in my data. (For further support of this point, see the impressionistic transcriptions below, as well as Figure 8.2 and the related discussion.)

---

5In fact, 56% of non-native-born Tuscaloosans in 1850 were from the Carolinas or Virginia (Foley 1972:3)
With the long-low-front slot of RA /aː/ considered to be empty (following Foley and McDavid), it might also be reasonable to reanalyse Southern monophthongized /ay/ (PRICE) as having shifted into that slot, since it is often phonetically long, monophthongal, low, and front. However, in Foley’s data, the sound corresponding to RA /ay/ is invariably a front-raising diphthong, which suggests that it belongs in the Vy subsystem (where it is located in Table 8.1).

This class of vowels, however, is quite commonly a monophthong, [a:], in Southern varieties. If this form were taken as fundamental instead of the also-occurring form [a*], then it would make sense to consider it to be a low, front, long monophthong and to write it as /aː/ rather than /ay/. This slot in the V: subsystem of Reference American was vacated by the assumed merger of RA /aː/ (broad A, or the PALM lexical set) with /a/ (short O, or the LOT set), so that the phonological prerequisite for a sound shift from low-front Vy to low-front V: is present: the target slot is empty. This may be part of the phonological motivation for the original monophthongization process; one analysis might put this vowel class simultaneously in both classes, while it shifts gradually from one to the next.

It should be pointed out that monophthongization of /ay/ is, or was, a conscious sociolinguistic variable (as is typical of late stages of long-established sound changes, before they go to completion, see Labov 1982)\(^6\) and the speakers Foley interviewed were undoubtedly using their most formal and self-conscious style, which is therefore likely to result in more gliding forms for /ay/. Thus it may still turn out that in this dialect the RA /ay/ class should be considered a long low-front monophthong at the post-lexical phonological level.

Unlike the relationship between RA /ay/ and /a/, where the nuclei are phonetically and phonologically identical, the nucleus of /ay/ is quite distinct from that of /a/ in this dialect. Unlike the realization of Reference American /oy/, which has a mid-back nucleus, Foley transcribes words of this class with phonetically lower-mid vowels. His transcriptions may be summarized as [ɔ(ː)⁷]\(^7\). These two facts suggest that these forms should be

---

\(^6\)This is, or was, a Southern shibboleth: monophthongs in the phrases “nice white rice”, or “bright light tonight” are stigmatized (Feagin, p.c., C.-J. Bailey, p.c.). Thus, speakers that monophthongize /ay/ before voiceless obstruents are marked as relatively lower-class speakers, though all speakers monophthongize in some other environments (C.-J. Bailey 1989:171).

\(^7\)Here as in SPE, () marks an optional element, and {} marks a choice among elements. It may be
categorized as /ay, oy/: low-front and low-back up-gliding diphthongs. These four vowels thus form a symmetrical set of low upgliding diphthongs, with front and back nuclei and glides:

Table 8.2: Low Upgliding Vowels in Alabama English.

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>Glide</th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>front</td>
<td>ay</td>
<td>oy</td>
</tr>
<tr>
<td></td>
<td>back</td>
<td>aw</td>
<td>ow</td>
</tr>
</tbody>
</table>

With the assumed merger of RA /a:/ (broad A, PALM) with /a/ (short O, LOT), and the shift of RA /o:/ (THOUGHT, CLOTH) to the Vw subsystem, there remain no long low vowels in this vowel system. This brings out a complementary distribution of the sound-classes within the structure: none of the V: vowels are low, while all of the Vy, Vw vowels are low. We might improve the elegance of the analysis by collapsing the long vowels into the Vw class, as in the base-5 system discussed in Phonological Preliminaries. The difference here is that the system remains base-5, and there are appropriate places in the system for the RA low vowels /æ, a, a:, o:/ (namely as Alabama /æ, a, a, and ow/, respectively), as in Table 8.1 above.

The Vr subsystem

Next, consider the vowels before /r/. Alabama speech is mixed between r-less and r-ful. Thus, the glide can be realized as [ɔŋ, ɔːŋ, ː]. It is symbolized abstractly here, as /r/.

The phonetic and social conditioning of rhoticity in this dialect is extensively studied by Feagin (1990). In most cases, while the sound symbolized by /r/ is not necessarily rhotic, it is distinct from other glides (except in one case, the confusion of /ɔr/ with /ɔy/ as [ɔŋ - ɔŋ], according to Foley 1972:36).

The inventory of r-gliding vowels is quite different from other dialects. This subsystem may be analysed as containing four heights of vowels in both the back and the front. This pointed out that this use of () notation is incompatible with C.-J. Bailey's use of () within phonetic brackets, [(X)..], which signifies a particular kind of systematic "optionality", namely that X is present in monitored styles, and absent in unmonitored styles.

See also McDavid (1948) for a very early sociolinguistic study of /r/ in another Southern variety.

225

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
would present a counterexample to theories of English phonological structure like the one presented in this thesis, in which only three heights may be represented. This thesis has found in all other cases that only three heights are necessary to represent English vowel structure, so an attempt is made here to reconcile this quite strong position with the facts of the Alabama Vr subsystem.

I will present the relevant contrasts first, and then show how another plausible analysis removes vowels from this subsystem, allowing the Vr subsystem to be analysed as having only three heights. First, it should be pointed out that the NURSE class, with a “pure” rhotic vowel /ɔː/ may be analysed as in Phonological Preliminaries as a vowel containing no underlying Nucleus, but only a rhotic (or [low]) Glide. The Nucleus is then inserted by a rule that ensures the well-formedness of syllables, and r-colored, perhaps by “backward-gemination” (C.-J. Bailey 1985). With the NURSE class excluded, the question now posed is, What are the underlyingly specified nuclei preceding tautosyllabic /ɪ/?

Alabama English appears to have four back vowels before /r/. First, the lexical sets CURE and FORCE are distinct here, as in Chicago and L.A. Chicano, but unlike Jamaican or Rhode Island, or possibly Appalachian English. However, unlike Chicago or Los Angeles, and like Jamaican, the sets, FORCE and NORTH, remain distinct, in this dialect, with such minimal pairs as morning/mourning, horse/hoarse, or/ore, war/wore, etc. (the first of each pair is in the NORTH class). Finally, there is a low-back unrounded nucleus in the START set, which remains distinct from the above sets. The back vowels are written as /ur, or, or, ar/, with corresponding lexical sets CURE, FORCE, NORTH, START.

The front vowels before /ɪ/ may be analysed in various ways as having from three to five heights. I will consider two contexts separately, before intervocalic /ɪ/ and before

---

9Steven Pelouquin, an undergraduate linguistics student at the University of Pennsylvania, has discovered a merger of chair and cheer, bear and beer, hair and here, etc., in his Rhode Island dialect. I have found two older Rhode Islanders that do not share the merger, according to minimal pair tests. But another one, a younger man, seems to share the merger, according to minimal pair tests, and furthermore merges the corresponding back vowels, the sets CURE and FORCE (also NORTH), so that bore and boor, tore and tour, etc., sound the same.

10Cf. Foley (1972:47, #13) and Kenyon, American Pronunciation, 10th ed., pp366-372, among many references. For this reason the Northern dialects which neutralize this distinction cannot be used as a starting point for the description of Southern dialects: it would be impossible to derive the membership of the FORCE and NORTH classes in a dialect that distinguishes them from a dialect which merges them. Wells' lexical sets are crucial in this context because they can be used to define most of the relevant sound classes in all English dialects, including this contrast in Alabama speech.
Before intervocalic /r/ there are as many as five heights: /irV, erV, ærV, arV/ are distinguished in weary, Mary, merry, marry, wiry. We may reduce the number of vowels that must be represented in the Vr subsystem to four by considering the distinction between /erV/ and /ærV/. This troublesome distinction may be eliminated in one of two ways, by merger, or by reanalysis.

In Foley’s data, 6 lower-class whites and 3 younger black and upper-class white speakers use /e/ in dairy and Mary, thus apparently merging /er/ with /ær/ (that is, with the sound in cherry, merry) in prevocalic position. This social distribution is reminiscent of a merger in progress, led by the socially intermediate group of non-cultured whites. For those who distinguish the Mary, merry word classes, the proposal made in Phonological Preliminaries must be extended to this dialect. There the Mary class of words was excluded from the Vr subsystem by supposing that there is a syllable boundary (written “$”) before the /r/, as in /ey$rV/. C.-J. Bailey (1985:167) writes Mary as [mə'ri] for older r-less Southern States speakers, a phonetic form which suggests that the /ey$r/ analysis of this class of vowels is correct, at least for those speakers. It may be that this syllable boundary is lost in younger and r-ful speakers, and that this loss is a trigger for a historical merger between /erV/ and /ærV/ (Mary/merry).

We may apply the same analysis to the /arV/ class. This vowel corresponds to /ay/ in other dialects, and in Phonological Preliminaries it was argued to be unnecessary in Trager and Bloch’s (1941) Vr vowel-system because the /r/ in Irish, spiral, wiry, etc., belongs to the following syllable. Thus we are left with three heights among the front vowels before intervocalic /r/: /irV, erV, ærV/.

Next consider the front vowels that occur before non-prevocalic /r/. There are again as many as five: ear [iər], Ayre [eər]: heir [eər]: air [æər] (assuming r-ful pronunciations). These correspond to the vowels before intervocalic /r/ and present similar apparent difficulties for the 3-height proposal.

Ayre, which rhymes with mayor and they’re, would seem to be /eər/ plus syllabic /ər/;

---

11 The merger of the Mary, merry classes occurs in many other dialects, including Los Angeles Chicano English, and Chicago White English, and Jamaican Creole.

12 C.-J. Bailey (1985:162) is the source of the Ayre/heir/air triplet. Another given there is they’re: their: there.
analyzing it as bisyllabic reduces the troublesome five by one.

*Heir* and *air* together correspond to the SQUARE lexical set. This distinction, between [e̞r] and [ɔr], seems to be a low-level rule, applying to words of the same class, rather than a distinction in the inventory of surface phonological classes. Thus in measurements of F1 and F2 in James H.'s speech, tokens of *there* are distributed across the entire range occupied by measurements of SQUARE words, and /ærV/ words are at both the high and low ends of the range. Thus *barrel* (n=1), *carry* (n=2) are relatively high and adjacent to *mulberry* (n=1), while *married* (n=2) and *rarity* (n=1) which impressionistically contain [æ] are low on the chart and adjacent to a token of *there*. *there* (n=23) is itself distributed from highest to lowest positions on the chart, which is consistent with the treatment of this difference as a variable phonetic alternation.

A difference does exist in James J.'s speech between *there* and *their*, but it is not the one suggested: *their* sounds like [ðɛə] (n=2), while *there* varies among [e̞r - ɛr - ɔr]. This may be derivative of the syntactic positions of the words: the determiner *their* is in a prosodically weaker position, and thus undergoes assimilation of the nucleus to the /-r/ glide.

C.-J. Bailey suggests that in certain contexts the /ær/ class is modified in a similar way by the rule of “backward gemination”, in which the /r/ quality bleeds back into the onset of the preceding syllable, and the nucleus consequently raises, giving /ær/ → [e̞r]. This is consistent with the information I have, but James H.'s two relevant tokens have undergone backward gemination to an extreme, with the nucleus becoming entirely rhotic.

It seems unnecessary to add a distinction between phonological classes /ɛr/ vs. /ær/, if the alternation between them can be accounted for by a low-level phonetic rule. So I will assume there is a single phonological class including all the SQUARE vowels, and write them as /ær/, reflecting the traditional Southern States transcription.

The argument so far has reduced the front vowels before /r/ to just three. But the four back vowels remain problematic for the three-height theory.

Three further points are necessary to complete the argument that the Vr subsystem can be analysed as having only three heights. The point of these steps is to enable the back vowel contrasts to be reduced by one.

First, /ær/, as in SQUARE, can be analysed as a phonologically mid-front vowel.
Second, /ar/, as in tire, can be analysed as similar to the same set of words in Reference American, as /ay$\$r/. Third, /ar/, as in the set, START, can be analysed as a front vowel. If these conclusions are accepted, then the Vr subsystem will have a three-heights-and-backness structure which is consistent with the theory of English vowel structure proposed in this thesis: there are only three phonological vowel heights.

Consider the points in order. The SQUARE set contains a mid vowel in Reference American and all the other dialects analysed in this thesis. As pointed out above, the pronunciation of this vowel with [æ] is by no means consistent in this dialect. Foley's reports that in addition to the predominance of [æ] pronunciations, there are a large number of mid pronunciations in words of this class, which he classifies as belonging to the [e] class. /ar/ in SQUARE does not occur in some related Southern dialects: in the self-conscious speech of a few emigrant Southerners from Texas, Arkansas, and Georgia, I have found a mid-front nucleus in SQUARE words. Finally, and most importantly in the present context, the speaker analysed here frequently produces a mid nucleus in these words, ranging from [ɛ] to [e], rarely as low as [æ]. (Even /arV/-class words like carry, barrel, have [ɛ] in his speech, though rarity and married, with impressionistic [æ], are exceptions). Thus the phonetic form of this class of vowel is perhaps more likely to be mid than low in the dialect under consideration. Certainly the case for analysing the SQUARE set as containing a phonologically mid nucleus is just as strong if not stronger than the case for analysing it in the traditional way, as low.

The second point can be posed as a clear question: Is the coda of tire /ar/, /ar/ or /ay$\$r/?13 Some rural speakers, such as James II., analysed here, appear to rhyme wire and (gui)tar, thus merging this class with the START class, here written /ar/. This may be a rural feature. /ay/ is monophthongized categorically in the words fire, tire, wire, etc. The sound is said to have a low-front nucleus, as [a:(h)]+/r/ (although James II.'s single token of this class, wire, is not front, but central: [wa$^{3}$]). This vowel is lower than /ar/ (SQUARE) and presumably in general in the South, if not in James II's speech, it is to

---

13Here, $ represents a syllable boundary, and /y/ represents not phonetic [y] but an abstract feature that is sometimes realized by up-gliding and sometimes by length. Surface phonetic [ay] sequences do occur, from phonological /a$\$y/ sequences. But while onset /y/ may surface as [y], coda /y/ does not. This is illustrated by the (r-less) Southern minimal pair, Maya [m$^{1}$ay$^{3}$], vs. Myer [ma$^{3}$]. pointed out to me by C.-J. Bailey (see, for example, 1985:118).
the front of /ar/ (START), and thus can be analysed as the low-front nucleus before /r/.

Where wire, tire, etc., are not merged with the START class, there are some reasons to analyse this complex vowel as bisyllabic /ay$ r/.

First, some pronunciations are bisyllabic (Foley, 1972:36), although these words may also be pronounced as a single syllable, even in highly-monitored speech. Second, /ay/ may still be analysed as an underlying diphthong, for a couple of reasons. Foley's transcriptions of /ay/ include no monophthongs.15

As discussed above, although all speakers monophthongize in certain environments (see C.-J. Bailey 1980:171), the monophthongization of /ay/, when it occurs before voiceless obstruents, is (or was) a stigmatized variable. This might be taken to suggest that the monophthongal form is not yet the underlying form. Further, while the monophthongal form can be derived from the diphthong by glide-deletion, it would be impossible to derive high-front-gliding forms in those cases that do have these glides, without an underlying feature that specifies that a high-front glide is to be added; but this amounts to having an underlying /-y/ glide as part of the representation for this class. Therefore, despite the frequent (in some environments, categorical) occurrence of the stereotypically Southern phonetic form [aː], we must analyse this vowel in this community as an underlying diphthong, /ay/, rather than as a monophthong /aː/ (though this seems to be undergoing phonological change at this time). Third, the nucleus of tire, wire, etc. (where these have distinct front [a] rather than central [a], merging with /ar/) is both phonetically similar and historically identical to the /ay/ sound. These three reasons suggest plausibly (though they do not make for certainty) that tire, wire, etc., may be phonologically diphthongal, and bisyllabic, written as /ay$ r/. If so, then since vowel sub-systems contain only the contrasts among vowels within single syllables, this sound-class does not belong in the Vr subsystem. Similarly words like spiral, with /r/ in the onset of the following syllable may be analysed as having an intervening syllable boundary and thus need not be included in this subsystem. Thus either through merger or a bisyllabic analysis, the tire, wire words need not be distinguished within this Vr sub-system.

To complete the 3-height analysis for Vr vowels, the last point is that /ar/, as in START, can be analysed as a front vowel. Acoustic and phonetic vowel space is an inverted triangle,
with [a] in the bottom corner. This is the location of the nucleus of /ar/. Since this point is at the intersection of the front and back edges of the phonetic vowel space, it is both front and back at the same time. The phonetic form doesn’t determine the phonological treatment of the vowel at the low corner of the triangle. So analysing it as phonologically front or back is an arbitrary decision with respect to the phonetics, which may be made according to the phonologically simpler structure. Thus if the phonology has a gap in the low-front slot and there is no space among the back vowels, it is justifiable to analyse /a/ as structurally a front vowel.16

To summarize, we may reasonably suppose that the Vr subsystem contains the following classes:

<table>
<thead>
<tr>
<th>front</th>
<th>back</th>
<th>front</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>ir</td>
<td>high</td>
<td>NEAR</td>
</tr>
<tr>
<td>mid</td>
<td>ær</td>
<td>mid</td>
<td>SQUARE</td>
</tr>
<tr>
<td>low</td>
<td>or</td>
<td>low</td>
<td>START</td>
</tr>
</tbody>
</table>

/ar/ is analysed as a [rhotic] Glide without a Nucleus specification, so it is also excluded from this subsystem, which strictly includes only vowels of the form N+G[rhotic] (that is, vowels with an underlying Nucleus position, followed within the same syllable by a [rhotic] Glide). Tire, spiral are not distinguished within this subsystem, being analysed either as bisyllables with the /r/ in the following syllable, or as merged with START. The vowels before intervocalic /r/ in the classes weary, merry, and marry/rarity correspond to the three front-vowels /ir, er, ar/, while Mary (=dairy), if not merged with the merry class, is analysed as having a syllable boundary before the /r/.

I have shown that the r-gliding vowels in this dialect may be analysed as retaining a 3-way height distinction among both front and back vowels. Thus despite the differences from Reference American in the lexical distribution of sounds (namely that FORCE and NORTH constitute two distinct phonological classes), and despite the important phonetic differences (e.g., the optionally lowered pronunciation of the vowel in the lexical set, SQUARE), the same static phonological structure of three heights and backness is retained.

---

16 In r-less Tidewater Southern speech (as in Boston), the /ar/ class is pronounced with a low-fronted vowel [a:].
The phonetic form of the /r/-glide itself does not rearrange this abstract phonological structure, though the phonetics needs to be fleshed out considerably. "R-less" and "r-ful" refers in most cases not to the presence or absence of a distinct phonological glide element, but to the phonetic form of that glide. This glide, which is abstractly characterized as /r/, varies impressionistically among [ɔ̃, ɔ̃, ɨ, ɛ] (Foley 1972:34), where the "r-less" forms [ɔ̃, ɨ] are "usual in socially prominent white speech and Negro speech, [and] occasional in other older white speech",17 and the "r-ful" form [ɔ̃] occurs elsewhere (p. 47, 48). These results are consistent with Feagin's (1990) interpretation that r-lessness is itself being lost in this Southern dialect. Varying realizations of /r/ are present in all Vr classes, excepting START words, which may have monophthongal [u:] for r-less speakers (p. 33). If we analyse the r-less and r-ful forms that co-occur in this dialect as derived from a single phonological system, then START-class words must have underlying /r/, whether realized as [ɔ̃] or [u:]. Thus the structural analysis may be retained, fleshed out by the particular phonetic and social distributions just presented, and those described in the discussion of acoustical measurements of the vowel system of a single older, rural, male speaker from Anniston.

8.3 The Shape of Vowel Space

The overall shape of James II.'s vowel space, shown in Figure 8.1, is unlike that of any other speaker studied so far. It appears as if the bottom of the vowel triangle was cut off, so that very little remains in the low corner. The entire top edge of the vowel space is very sparsely populated, unlike Chicano and Chicago, where the top edge of vowel space is the most dense part of the entire distribution. Unlike the other dialects, there is no gap in the high-central area relative to the high-front and high-back regions. The filled gap may be due to the fronting of /u:, u/. The generalization that there is always a gap in the high-central region of vowel space appears to be incorrect. Instead, it appears that phonetic vowel systems may or may not make use of that region of vowel space.

17According to Foley (1972:50), [ɨ] is more restricted than [ɔ̃], occurring in /ɔr, or/ words for blacks, but only in /or/ words for whites.
Figure 8.1: 1637 measurements of F1, F2.

Alabama Vowel Space: James H., 80
Table 8.3: Impressionistic transcriptions of stressed vowels

- /i/: [ɪ, ɪ, ɪ', ɪ², ɪ⁹, ɪ⁵, ɪ, ɪ', i'/ɪə, ɪ, ɪ', ɪr, ɪ', ɪ' /ɪə, ɪ, ɪ' /ɪə, ɪ]  
- /u/: [u, u, u, u, u, u, u, u]  
- /ɛ/: [u, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ]  
- /ʌ/: [ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ]  
- /æ/: [ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ, ɛ]  
- /ɑ/: [ɑ, ɑ, ɑ, ɑ, ɑ, ɑ, ɑ, ɑ, ɑ, ɑ]  

/i:/ [ɪ', ɪ', ɪ', ɪ', ɪ', ɪ', ɪ', ɪ', ɪ', ɪ', ɪ', ɪ']  
/ey/: [e', e', e', e', e', e', e', e', e', e', e', e']  
/ow/: [ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ, ʌ]  
/əw/: [ə, ə, ə, ə, ə, ə, ə, ə, ə, ə, ə, ə]  
/ə/w/: [ə, ə, ə, ə, ə, ə, ə, ə, ə, ə, ə, ə]  

8.4 Impressionistic Transcriptions of the Stressed Vowels

The stressed tokens of all non-rhotic vowels were listened to 3 times each and transcribed impressionistically. The distinct phonetic characterizations found are presented in Table 8.3.

These transcriptions show a number of important phonetic facts about this Alabama dialect, some of which will be pointed out here, and some discussed in conjunction with the acoustic measurements discussed below. The nuclei of /uw, u, ow/ (Wells' GOOSE, FOOT, GOAT sets) are very fronted, especially /uw/. /ay/ is variably monophthongized, unlike in Foley's data. The nucleus of /aw/ is fronted and raised to varying degrees, and its offglide is sometimes a back or even back-and-downward glide, as noticed elsewhere. /ɔ/ is often a mid-back diphthong with a high-back glide. The single token of /oy/ is consistent with an analysis of /oy/ as a two-syllable sequence beginning with /ɔ/. The short vowels /i, e/ and sometimes /ʌ, æ/ have inglides.

This is one of the main phonetic facts of this dialect: the ingliding of short vowels /i, e, ʌ, u, æ, ɑ/ in stressed monosyllables, documented for all six vowels separately by Foley. Inglinging among the front vowels alone is discussed in C.-J. Bailey (1969). The breaking and gliding of /æ/ to [æ, æ, æ] is discussed in Feagin (in press).
What is the correct formal phonological treatment of this ingliding pattern? One rather abstract analysis that might be proposed is an extension of the Reference American principle, "Stressed rhymes branch" (discussed above on page 65). In this variety, the related principle is:

In monosyllables, stressed nuclei branch.

In order to satisfy this principle, short vowels in monosyllables must have an extra timing slot added to them, which may be filled in as the null vowel, or schwa, thereby deriving an inglide on short vowels in stressed monosyllables.

8.5 Instrumental Evidence for Sound Change.

This section will present and discuss estimates of means of distributions in formant-space of this speaker's vowels. For each vowel, the estimated range of possible values for the mean of the vowel's F1-F2 distribution was calculated by using the bootstrap technique. The discussion of the bootstrap technique in the Methods chapter is prerequisite to understanding these distributions. 200 "bootstrap" resamplings were done of each vowel class, and the mean for each resampling was plotted. The distribution of resampled means gives a near-optimal estimate of the range within which the true mean lies. Of course these clouds do not represent the distribution of all the tokens of the particular sound-classes, but only of estimates for the mean of that distribution. The sound-classes themselves overlap considerably, presenting a problem of classification which may or may not be completely solved by additional information such as formant-trajectories, duration, stress level, and phonetic context.

Figure 8.2 shows the distributions of bootstrapped means for 13 non-rhotic Alabama vowels in F1-F2 space, as estimated from 1637 measurements of vowels in the conversational speech of one rural, working class, older white male speaker. Each cloud of points represents the location and inherent accuracy of the estimated mean for the given vowel class, using the particular set of measurements made here. The way to interpret each distribution is that the true mean of the vowel class (assuming a similarly constructed sample whose size increases towards infinity) could be anywhere within the cloud with equal likelihood.
Figure 8.2: Bootstrapped mean F1, F2 for each vowel class.

AE (James H.): 200 Re-estimated means for each vowel
F1 is a measure of the degree of mouth-opening, while F2 reflects the degree of tongue-body frontness (Chapter 2, Acoustics, justifies these statements), which are the primary phonetic parameters distinguishing vowel quality. Thus the chart is an acoustic representation of vowel quality at the nucleus measurement point within the syllable. The origin (0,0) of the axes is to the upper right of the charts; this reverse orientation, created by plotting the negatives of the measurements, makes this graph correspond quite precisely with the usual auditory chart and articulatory vowel-triangles: high-front is located on the upper left, low central is in the middle on the lower edge of the graph. This representation shows in fine detail many differences in phonetic quality among different vowel sounds.

In comparing the locations of the vowels for this Alabama speaker with those that may be expected for Reference American or with those found in other dialects, a number of striking vowel shifts can be observed. The back vowels undergo a number of changes. First, /uw/ shifts very far to the front, so that it is adjacent to /iy/. /u/ also fronts quite far, especially in unstressed syllables. /u/ is farther front than /ʌ/. These facts confirm the impressionistic results presented above.

The distribution of bootstrapped means for /u/ is much more widely dispersed than the those of other vowels, because there are only 18 tokens of /u/ in the data for this speaker. In fact, the generally more dispersed clouds in this data as compared with, say, the primary Jamaican speaker (Juba B.), is due to the smaller number of measurements (1637 as opposed to 2680).

/ow/, the mid-back long vowel, may undergo some raising, and rather more fronting, judging from the impressionistic transcriptions above. /ow/ is less peripheral than the THOUGHT-class vowel, /ɔw/, which has become a back-raising diphthong, and is quite high; it is certainly not phonetically a low-back vowel, but more like [oː]. With this form, the contrast between /ow/ and /ɔw/ for this speaker is based on a very small phonetic difference, as can be seen by the fact that they are immediately adjacent to each other in Figure 8.2. These vowels, which share the phonetic feature of a high-back glide, and whose nuclei are distinguished only by degree of peripherality in the mid-back region, may form a near-merger, an interesting matter for future perceptual research in this dialect. This approximation of the two classes is quite different from the expected phonetic contrast for younger or more upper-class speakers documented by Feagin, Foley, C.-J. Bailey, and
others. For example, Feagin (1990:143) gives vowel charts for three other (upper-class or young or both) Anniston speakers, where all three clearly have low nuclei for /aw/ and mid or upper-mid nuclei for /ow/. The expected contrast might be written as [ʌ⁰] versus [ɔ⁰]. Further work would be necessary to determine the social and linguistic patterning of these interesting and mysterious differences.

/α/ is shifted up and back from the nucleus of /ay/, which it is identified with in dialects like Reference American. The back-round status of the LOT and THOUGHT classes (/a, ɔw/) comes from one of two sources. It may be a historical relic: these vowels are the long and short mid, back, round vowels of earlier forms of English, and in this dialect they may not have followed the lead of other American dialects in the lowering of LOT from [o] to [a] and of THOUGHT from [ɔ:] to [ɔ:]. The other hypothesis is that they had previously lowered and fronted, and now have undergone a retrograde shift; this is somewhat less plausible, since it requires a greater number of changes.

Another long low vowel, /aw/ (MOUTH), undergoes a change that may be related to the diphthongization of /ɔw/ (THOUGHT). When impressionistically transcribed, the nucleus of /aw/ is found fronted and variably raised to the area of [æ, e]. If /ɔw/ moves from the long-vowel subsystem into the back-upgliding subsystem, then it becomes sensible that the nuclei of the two vowels should move apart so as to maximize the phonetic difference between the two. This may explain both the fronting of /aw/ [æ⁰] and the raising of /ɔw/ [ɔ⁰]. Additionally, the glide of /aw/ is not always an up-glide; sometimes it glides directly back, and even back and down from the nucleus, as in [e³] (as noticed in Philadelphia by Labov, and documented widely through the South by G. Bailey, and colleagues).

A smaller number of shifts operate on the front vowels. /æ/ raises nearly to mid (this may be an effect of the general close-mouthed character of James H.’s speech, though, since /æ/ is the lowest of the front vowels), and often has high-front glides. /eə/ shifts in the opposite direction, falling and laxing so that it overlaps with /ɛ/. This overlap of the nuclei of /eə, ɛ/ does not result in homophony, since glides distinguish the two.

All of these vowel shifts obey the rules for vowel-shifting in Labov, Yaeger and Steiner (1972), mentioned also in other chapters: In chain shifts, back vowels front, peripheral (long) vowels raise, and non-peripheral (short) vowels lower. Thus back vowels /uw, u/, and, to a slight extent, /ow/ move to the front, short /a/ falls to a low position, and long
/ɔw/ raises along with /æ/.

Additionally, nucleus-glide differentiation applies to both /aw/ and /ey/ to front the nucleus of /aw/ away from the back and to centralize and then lower the nucleus of /ey/ away from the high-front-peripheral location of its offglide. /ey/ is lowered and centralized so far that its nucleus is indistinguishable in F1-F2 space from that of /ɛ/, as shown in Figure 8.2.

8.5.1 Nucleus-Glide Differentiation

The rule of nucleus-glide differentiation can be understood as assuming a fixed glide endpoint, and shifting the nucleus away from it. However in the shift of /aw/ to [eə], both nucleus and glide are shifting simultaneously, and in opposite directions. This understanding of nucleus-glide differentiation must be incorrect, if this case is an instance of it. When the glide falls to mid, and the nucleus raises to mid, the result is that both nucleus and glide are mid, whereas before they were low and high, respectively, and thus became less distinct.

The other case I know in which the glide shifts as well as the nucleus, is French “moi”, which derived from Latin /me/, presumably by diphthongization to /mei/, nucleus-glide differentiation to /maɪ/ and /moɪ/ (the stage reflected in the current spelling). The current form is /mwa/, showing three further changes from the /moɪ/ form: The first segment moves to high-back; syllabicity shifts from the first to the second segment; and the second segment falls from high to low position, resulting in /moɪ/ → /mwa/. The ordering of the last three shifts cannot be ascertained from the two-stage picture provided by the spelling and the current pronunciation.18

In both the shift of Southern /aw/ to [eə] and French /oi/ to [ua] (not specifying syllabicity here), the two segments in each vowel move in opposite directions, perhaps simultaneously. In the French case it is the front vowel that falls while in the Southern case it is the back vowel that falls; in both cases it is the second segment that falls and the first that rises. Why the both nucleus and glide should both shift simultaneously is a mystery.

---

18 C.-J. Bailey points out to me that Southern boy as a call to a dog, or a threat to a youth, can undergo the identical shift as French moi [bwa:], and similarly that goin’ may be rendered as guine.
Nucleus-glide differentiation may be related to the rules of sound change based on peripherality of Labov, Yaeger, and Steiner (1972), and to the vowel-before-vowel raising rule of Kenyon and Knott (1941:xxxviii, #92), discussed in Phonological Preliminaries. But an explanation for this shift of /aw/ to [e°] has yet to be found.

8.6 Stress and Vowel Quality

This section considers the effects of phrasal stress on vowel quality as represented by F1-F2 measurements. See Chapter 5 for a discussion of what stress means here (that is, how utterances were impressionistically coded for stress). Somewhat different effects are found when clitic words are included or excluded. Since the pattern is simpler, statistically more significant (4 more vowels have significant effects), more similar to the pattern found in other dialects, and more understandable when clitic items are included, I will primarily discuss the pattern found when all tokens are included, though both charts are displayed, for consistency with other chapters.

Consider Figure 8.3, which displays what happens to vowel nucleus formant frequencies under stress reduction.

The interpretation of the graphs is as follows. Each arrow is labelled with the symbol from Table 8.1 for the class it represents. The tail of each arrow is at the mean of the measurements of the stressed (primary and secondary together) tokens of that vowel class. The head of the arrow is at the mean of the unstressed tokens. Thus each arrow shows the average effect of destressing on the phonetic quality of that vowel. Solid arrows represent statistically significant differences between the stressed and unstressed sets for that class, using the two-tailed t-test described on page 153. Dashed arrows are insignificant at the 5% level on this test. Vowel classes which only occur with stress are plotted at their mean location and marked with ', while vowels that only occur without stress do not have the apostrophe. The clitics-excluded chart displays the effects of stress on vowel classes from which tokens occurring in clitic words are excluded, while the lower chart displays the effects of stress on vowels including both clitic and non-clitic tokens. The clitic/non-clitic distinction is defined in the Methods chapter.

Note that /a/ does not include tokens of the reduced vowel /i/ argued by Sledd (1966).
Figure 8.3: Effects of impressionistic phrasal stress on vowel nuclei in F1-F2 space.

AE(James H.): Effects of Destressing.
Clitics excluded

AE(James H.): Effects of Destressing
Clitics included

Dashed arrows are insignificant on 5%, 2-tail t-test

241

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
to be phonologically distinctive in Southern dialects. Instances of this class are classified as unstressed tokens of /i/, which is arguably the stressed vowel corresponding to /i/. The mean for this vowel class is thus located at the head of the arrow whose tail is labelled with “l”.

The pattern of reduction is clearer (more consistent with the patterns found in other dialects) in the clitics-included chart than in the clitics-excluded chart, so I will discuss the clitics-included chart. There we find that the front vowels raise and centralize, with increasing centralization from /æ/ to /ey/ to /ɛ, i/ and finally /iy/. The central vowels (fronted /u/ as well as /a, a/) move up and to the front. All of these are consistent with the theory of reduction as a shift in formant frequencies roughly in the direction of a target around [i], where [i] is understood as at least as high as [i]. This tendency is rather rough, since /i, e, ay/ point towards a location somewhat to the back of this target.

A second pattern is observed in this chart, namely the reduction of the nucleus of a diphthong in the direction of its offglide, or nucleus/glide assimilation. This effect explains the two anomalous, though statistically weak, shifts of /ow/ and /aw/.

/ow/ has a high-back glide, and when destressed it shifts toward the high-back corner. Similarly /aw/ frequently has a low-back glide as [ə], and when destressed it shifts back and slightly downward. Nucleus/glide assimilation might also be used to explain the anomalous direction of shift for /i, e, ay, ow/. /i, e/ are inglides in stressed monosyllables, according to Foley (1972); these vowels centralize more than they would if they were simply shifting in the direction of the apparent reduction target [i]. /ay/ is impressionistically transcribed above as having central rather than front glides in some cases, and the stress-reduction effect on /ay/ does not include any fronting. If /ay/ simply shifted toward the reduction target when unstressed, it would shift to the front as well as upwards; if instead /ay/ shifts towards its high-central glides, it would not shift to the front — this is the actual effect found.

/ow/ also does not shift towards high-front-central position. /ow/ has a high-back glide, and under stress-reduction it moves — just slightly — toward the high-back corner.

It is important to note that the reduction target, [i] seems to be quite distinct from the location of the reduced vowel, /a/. The reduction target is high and well to the front of central, while /a/ is upper-mid, central, at about the same location as the mean of
stressed /u/ and unstressed /ɔ/ (/ɔ/ is presumably distinguished by F3 differences). The reduction target for this dialect appears to be high and relatively front. It is different, then, from the mid-central reduction target of Jamaican as well as the high central reduction target of the Chicago vernacular.

8.7 Summary

This chapter has discussed the vowel system of a single speaker from Anniston Alabama from a number of directions. The surface inventory of the Alabama vowel system was explored at length and a proposal for its phonological structure was made. The overall shape of vowel formant space was examined. Impressionistic transcriptions of stressed tokens of each of the vowels were listed and discussed. Bootstrap resampling was used to display estimates of the mean locations for each vowel. These formant-frequency means were described in terms of historical sound changes that influenced them. The effects of stress on vowel quality were explored, finding effects of reduction towards a high, somewhat front reduction target, as well as effects of reduction of nuclei in the direction of their offglides. Differences in the effects of following consonants between this speaker and those of other dialects are discussed in Chapter 10, where it is shown that vowels occurring before /ŋ/ are significantly lowered, a phonetic rule which seems to be characteristic of this dialect. While much remains in the description of the linguistic system of phonetic implementation for this dialect, some important steps have here been made in that direction.
Chapter 9

Los Angeles Chicano English

This chapter discusses the phonology and phonetics of the ethnic dialect of the Chicano (Mexican-American) community in Los Angeles, California. The inventory and structure of the (surface) phonological system are discussed. Phonetic patterns are displayed in charts of acoustic measurements which reflect the processes of phonetic grammar. These processes or relationships are discussed below; they specify the mean location (one might say, target) of vowel classes in vowel space, and characterize the significant shifts from these locations that are correlated with phrasal stress and with preceding and following consonants. Some evidence for the low-back merger, between Reference American /a/ and /a:/ (as in LOT and THOUGHT) is also examined

9.1 The Community, the Individual.

An explosion of Mexican immigration\(^1\) began in this century with the exodus of refugees from the Mexican Revolution (1910) and the linkage of Mexican railroads to the U.S.(Santa Ana, 1991). The Hispanic population is one of the largest and fastest-growing ethnic groups in America. In the Los Angeles area alone, they form 40% of the population (roughly 1.4 out of 3.5 million, in the 1990 Census). The result of this migration, and the segregated

---

\(^1\)The resident Mexicans, or "Californios", who remained in California after the Mexican war in which California was taken away from Mexico, are a tiny population, mostly assimilated to the Anglo community. The entire recorded population of Los Angeles in 1820 was 615, and of California at the time of the 1848 war, about 15,000. The populations discussed here, both Chicanos and Anglos, is entirely composed of later immigrants.
social conditions the immigrants found in California, is an ethnic community that is only partly assimilated to the matrix “Anglo” (that is, European-American) community. It retains symbolic links with Hispanic culture (as well as real links through continuing immigration), but linguistically is mostly an English-speaking rather than a Spanish-speaking community, though its members have a distinctive accent. The phonological inventory appears to be identical to that of the local Anglo community. For example, the long and short vowels (/i:/ vs. /ɪ/ , /u:/ vs. /ʌ/, etc.) are clearly distinguished, as are the relatively rare English vowel classes /æ/ and /ɑ/ (for confirming evidence, see Figure 9.2). Speculatively, it seems that the main differences between the Chicano accent and the local Anglo accent are first, that the Chicanos are not participating in the ongoing phonetic changes in the Anglo communities (the raising of /æ/); and second, that there are distinctively Chicano elements in the prosodic system. Santa Ana (1991) discusses phonetic variation across the Chicano community.

The sole subject of this chapter’s phonetic investigation of the Los Angeles Chicano English dialect (henceforth LACE) is a 30-year-old working class Chicano male who we will call Vince. While Vince claims to have spoken Spanish natively as a child, and may have some passive Spanish competence, he is now a monolingual speaker of English. Santa Ana observes that he cannot speak even a few words or phrases of Spanish, though Spanish does carry considerable symbolic significance for him. Vince’s mother is a native English speaker, as was her father (according to his statements), so he is a third-generation English-speaking American.

This reiterates the point that the Chicano community is not, as may be naively supposed, a Spanish-speaking community of second-language English learners. While immigration of Spanish speakers from Latin America continues unabated, the great majority of the children of immigrants in the barrios (Chicano neighborhoods) of the Los Angeles area are monolinguals, including Vince and his age-cohort. Thus Vince grew up a native

---

2 As in FLEEECE, KIT, GOOSE, FOOT, respectively.
3 As in TRAP, NURSE.
4 The issue here is whether or not there was exposure to natively-spoken English in the household during childhood. The first person in the family with native competence is the first-generation speaker, in this view. All those in the family with a native-English speaking parent are exposed during childhood to natively spoken English, whether or not there is also non-native English in the household, spoken by family members that are immigrants. Thus if even a single grandparent is a native English speaker, the grandchild of such a speaker is third-generation.
speaker among native speakers.

Vince lived in East Los Angeles until the age of six, and afterwards in South San Gabriel, a town even farther to the east. The typical pattern of settlement for Mexican immigrants is to initially live in a “port of entry”, such as East Los Angeles. Mexicans and Chicanos find it preferable to maintain frequent (daily, or at minimum weekly) contact with kin. Since nearby housing enables more frequent contacts, the most desirable place to live is next door to one’s parents. This preference for close familial ties, in addition to limited housing opportunities, and the constant influx of new immigrants results in a very crowded housing situation, since there are only so many places close to the ports of entry. When children of immigrants grow up, then, they look for the nearest available housing — which is no longer in crowded East L.A. Instead they move, typically towards the east (away from the ocean, into less crowded and less expensive housing areas), to nearby towns where housing is available. Thus the lower-working-class neighborhood where Vince was raised is not coincidentally in the town of South San Gabriel, a few miles east of East Los Angeles. This migration pattern is continuing, as the area gets increasingly crowded, with people moving as far as Riverside (50 miles away) or Antelope Valley (70 miles away). Vince grew up in a neighborhood in South San Gabriel that was a mixed Anglo and Chicano area in his youth, but which has gradually lost whites and gained Hispanics (as well as Asians) over the last 15 years.

Vince is a respected member of the social group to which he belongs, a circle of Chicano friends and associates that live in the same neighborhood. He is married, with children, and is settled in the community. This speaker was chosen as a core member of the English-speaking Chicano speech community, which in a single family has resided for generations in Southern California. The interview is one of a corpus of 150 interviews done by my colleague, Otto Santa Ana, as part of his dissertation research on the Los Angeles Chicano speech community. See also the description of Vince in Santa Ana (1991, Chapter 5).

5The population in 1930 of East L.A. was 70% of 1970’s population, whereas in West L.A. it was 3%.
6We have collaborated on the acoustic analysis of this and other speakers, working together to some extent since our interests are complementary. His work focuses on social variation within the Los Angeles Chicano English speech community, considering consonant deletion, and variation in the phonetic quality of stressed vowels, while my work focuses on the internal sources of variation in vowel quality, according to consonant environment and stress.
9.2 LACE Vowel Structure

What is the structure of the L.A. Chicano English vowel system? How does LACE compare to other dialects of English phonologically? What are the phonemes, the available contrasts? What are the word-classes associated with these classes and how do these compare with other dialects? This section answers these questions.

As in other chapters, I will use the lexical sets defined by J.C. Wells (1982) for comparison, pointing out the groups of lexical sets which are not phonologically distinguished in LACE. These are labelled “merged sets.”7 Splits will not be discussed, since I found no evidence for them.

Many American dialects do not distinguish the word classes NORTH and FORCE (though Southern dialects like that of Anniston, Alabama, do keep them separate). I assume that LACE is among these. Like other American Englishes, LACE is a “flat-BATH” dialect. That is, it classes the BATH set with the TRAP set rather than with the PALM set. A third issue is whether or not Chicano English distinguishes the low-back vowels /a/ (THOUGHT, CLOTH) and /a/ (LOT). This historical merger is progressing rapidly on the West Coast of the U.S.8 I will examine the question of whether or not LACE makes this distinction, but for now I assume it does not. Hence, for my purposes Los Angeles Chicano English does not distinguish the following lexical sets of J.C. Wells:

\[
/\text{a}/: \text{LOT} \sim \text{THOUGHT} \sim \text{CLOTH} \sim \text{PALM} \\
/\text{a}/: \text{TRAP} \sim \text{BATH} \\
/\text{or}/: \text{NORTH} \sim \text{FORCE} \\
/\text{or}/: \text{NURSE} \sim \text{LETTER} \\
/\text{i}/: \text{FLEECE} \sim \text{HAPPY}
\]

Because of phonetic similarity and complementary distribution, stressed and unstressed /\text{or}/ (NURSE, LETTER) are the same phonological class (unlike in Jamaican, where the

---

7Using the term “merger” is misleading because actual phonological mergers have not necessarily taken place in each case in this particular speech community. Nevertheless, this remains a convenient term when referring to lexical sets that are not distinguished.

8Thus, my parents, both born and raised on the West Coast, maintain the distinction, while I and my sisters do not. Previous studies of this merger are discussed below.
unstressed vowel in LETTER is not rhotic and is classified as /a/, while the stressed rhotic vowel in NURSE remains phonologically /ɛə/). Similarly, stressed and unstressed high-front-peripheral vowels (FLEECE, HAPPY) are classified together as /iː/ (unlike in older RP (cultivated Southern British), where HAPPY ends with the vowel in KIT).9

The phonological structure assumed here is presented graphically in Table 9.1.10 11

This structure is somewhat different from the Reference American structure (Table 3.10), primarily due to mergers that have occurred in the historical development of LACE. These mergers include the merger(s) of RA /a/, /aː/, and /ɔː/ (Wells' LOT, PALM, and THOUGHT sets, traditionally called “short O”, “broad A”, and “long open O”, respectively). While the merger of broad-A with the other vowels may have occurred some time ago, the “Low-Back Merger” of Labov (1991) between Reference American /a/ and /ɔː/ appears to be currently in progress in California Anglo English (see section 9.8). The resulting phonological class, here written as /a/, must be analyzed as a phonologically long vowel, given the basic English principle that short vowels cannot be stressed without


10The typography used, with /iː/, /i/, /uː/ as the high front and back vowels, is intended to avoid the ambiguity of /i/ and /u/ which arises when these 4 vowels are alternately symbolized with the sets /i, i, u, u/ or /iː, i, u, u/. Similarly, “e”, “o” are avoided due to this ambiguity. The point in phonological transcription is to be unambiguous.

11The lexical sets (of Wells, 1982) that correspond to these phonological classes are given below in order to clarify the correspondences of these classes with those of other dialects, including the reader's. See page 42ff for discussion of what lexical sets signify and how they are useful.

BATH is included with TRAP; FORCE is included with NORTH; PALM, CLOTH, and THOUGHT are included with LOT, Unstressed vowel classes HAPPY, COMMA, LETTER, and the /ɔː/ vowel, NURSE, are not shown here.

248
a following consonant in the syllable.\textsuperscript{12} LA Chicano /\textipa{a}/ occurs in such words as \textit{law}, \textit{spa}, etc. If this general English principle applies to this dialect, this sound cannot be analysed as a short vowel. Nor can it be analysed as a vowel with a glide. First, its low, non-front nucleus is identical with that of /\textipa{aw}, \textipa{ar}/ and possibly also /\textipa{ay}/. Second, it contains no phonetic glide. Thus it cannot be structurally analysed as occurring within the Vr, Vv, or Vw subsystems.

The remaining alternative is to place /\textipa{a}/ within the V: long vowel subsystem. If this were not the only alternative, then the other long vowels, /\textipa{i}:, e:, o:, u:/\textsuperscript{13} might be analysed phonologically as elements of the Vy and Vw systems, thus eliminating the V: subsystem, as in Alabama English (page 222). But there is no such alternative (barring modification of the fundamental English principle that stressed rhymes branch), and so the proposed structure (Table 9.1) of the LA Chicano vowel system locates /\textipa{i}:, e:, o:, u:/ and /\textipa{a}/ in the V: long vowel subsystem, and provides separate subsystems for the distinctive classes of glides that occur with /\textipa{a}/: /\textipa{ay}, \textipa{aw}, \textipa{ar}/. The large number of gaps in this system is due to the various mergers that have eliminated vowel classes from the structure, and to the need to posit Vy, Vw subsystems, which itself derives from the apparent identity of /\textipa{a}, \textipa{ay}, \textipa{aw}, \textipa{ar}/.

The non-high front vowels before intervocalic /\textipa{r}/ are presumably merged in this dialect (as in the local Anglo dialect and in Chicago, but not in Philadelphia, and various Eastern dialects). That is, \textit{Mary}, \textit{merry}, \textit{marry} are pronounced identically. This phonological collapse has two simplifying effects. First, it eliminates a rather tenuous distinction based on syllable structure rather than segmental features: \textit{Mary} and \textit{merry} are elsewhere distinguished phonologically as /\textipa{me:}$\textipa{ri}/ and /\textipa{mer}$\textipa{si}/.\textsuperscript{14} Second, by eliminating the /\textipa{arV}/ class by merger with /\textipa{erV}/, the last instance within LACE of a front-back contrast among low vowels is eliminated. There is only one low vowel within each of the four subsystems: /\textipa{ae}, a, ar, ay, aw/ are the sole low vowels for the V, V:, Vr, Vy, and Vw subsystems, respectively.

\textsuperscript{12}This is formalized on page 65 for Reference American: Stressed rhymes branch.
\textsuperscript{13}These vowels are sometimes written more phonetically, as they occur in other dialects: /\textipa{iy}, ey, ow, uw/. Both transcriptions can be accurate at the same time, depicting different stages of the derivation of the phonetic form. There is no problem of ambiguity among the classes referred to, and the particular transcription chosen is an arbitrary decision. However, in this dialect, these vowels are frequently monophthongal — probably an ethnic marker that the LA Anglos do not share.
\textsuperscript{14}Note that $\textipa{s}$ signifies a syllable boundary.
Since base-5 systems (like Jamaican Creole) containing one low vowel but two mid and two high vowels are phonetically simpler (though phonologically less symmetric) due to the triangular shape of phonetic vowel space, this elimination of the front-back contrast\textsuperscript{15} has simplified an aspect of the system. The proposal above could be modified so as to have a base-5 structure, as in Jamaican Creole, at the cost of an unnatural rule that fronts the nucleus of the low short vowel to /æ/. No perfect solution is apparent, and for now I will accept Table 9.1 as the surface-phonological structure of the Los Angeles Chicano English stressed vowel system. The same structure might also be applicable to the Anglo community.

### 9.3 Impressionistic Transcriptions of Stressed Vowels

Impressionistic transcription of long, stressed tokens of these vowels in monosyllabic words, excepting those before /r/, results in the range of impressionistic realizations of the above phonemes presented in Table 9.2.

Table 9.2: Impressionistic transcriptions of stressed vowels

<table>
<thead>
<tr>
<th>/i:/</th>
<th>/i, i', i:, i:]</th>
<th>/e:/</th>
<th>/e, e, e', e']</th>
<th>/\theta:/</th>
<th>/\theta, \theta', \theta:, \theta:]</th>
<th>/o:/</th>
<th>/o, o, o', o:]</th>
<th>/u:/</th>
<th>/u, u', u:]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>/a, a, \alpha, \alpha']</td>
<td>/\omega:/</td>
<td>/\omega, \omega', \omega:, \omega:]</td>
<td>/\omega/</td>
<td>/\omega', \omega, \omega', \omega:]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/\omega:/</td>
<td>/\omega, \omega', \omega:, \omega:]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some observations are worth noting. First, this speaker is not strongly participating in the widespread raising of /æ/, since he produces cardinal #4, [æ], in stressed tokens, though coarticulated tokens sound more raised, as far as [e].

In fact, the opposite change is occurring in one environment: /æ/ falls towards [e] before /l/, as in [æ]levator, Mr. B[æ]lvedere, etc. Since /æ/ does not exist in Spanish, the fall of /æl/ cannot be attributed to Spanish influence. On the contrary, the fall of /æl/ seems to be a purely English sound change that happens to occur in this particular ethnic group, not a contact- or contact-induced phenomenon.

\textsuperscript{15}For example, a near minimal pair would have been marry vs. sorry. The contrast between them here is not simply front vs. back, but also mid vs. low.
/u:/ is somewhat front, as in most American and many British dialects. Anglo speech in Southern California shows even greater fronting of /u:/, to such an extent that /u:/ and /u/ overlap with /i:/ and /ɪ/ in formant space.

According to Santa Ana (1991), even young Spanish-speaking immigrants who arrive in Southern California at 8-10 years of age, learn English with /u:/ pronounced so far to the front that it overlaps with /ɪ/. Tokens are high and central, just as in an unpublished analysis I have done of a California Anglo speaker. For all speakers Santa Ana studied, Chicanos front /u:/ to overlap with /ɪ/ (except one, whose /u:/ is as still as front as /ɪ/, but is higher).

/aw/ has some fronting of the nucleus and falling of the glide. In Philadelphia and many Southern dialects, the nucleus of /aw/ has shifted toward [æ] and sometimes goes as far as [e], while the offglide falls from [u] to [o] and sometimes as far as [ɔ].\(^\text{16}\) Vince thus appears to be participating in this shift of /aw/, but not to the extremes found in other dialects.

Finally, some realizations of /iː, eː, oː/ and other long vowels were transcribed as monophthongal. This may be an effect of Spanish, though other American dialects (Minnesota, and Wisconsin, for example) also show monophthongization of these vowels, which are most commonly diphthongs in English.\(^\text{17}\) Also, these vowels are underlyingly long monophthongs, so the general effect here is to simplify the system of phonetic implementation, as compared with the /iy, ey, ow, uw/ of many other English dialects.

These impressionistic transcriptions need some qualification: they give a rather qualitative idea of how one transcriber perceived a small number of tokens on one particular occasion. While the impressionistic transcription of a few stressed tokens can generate considerable insight, it is difficult to summarize large quantities of impressionistic data, and it may also be difficult to use such data to substantiate the finer allophonic differences which are the main focus of this research. Further, it is difficult to accurately transcribe short, coarticulated, unstressed tokens, which may behave differently from the easily transcribed long and stressed tokens. For these reasons, it is important to have more — and more consistent — data on vowel quality. If the distribution of a great many tokens in phonetic

\(^{16}\)This may have first been noticed by Labov for Philadelphian white vernacular, and has been shown to occur much more widely, especially throughout the American South, by Thomas, Bailey, & Benson, 1990.

\(^{17}\)See discussion of this point on page 100.
vowel space could be viewed at once and summarized, that would constitute a more trustworthy representation of vowel data. Such a quantitative and objective representation is possible using formant-frequency charts. As shown in Chapter 4, small but measurable differences in formant frequency can have clear, perceivable consequences. Therefore we may assume that much of the information on a chart of F1 vs. F2 frequency measurements reflects perceptible vowel quality. Thus we now move from impressions to instrumental measurements.

9.4 Overlap of Phonemes

As Lisker (1949) showed, distinct pronunciations of distinct phonemes overlap in formant space (in that case, /æ, e/, in the context, p-p.). As I showed above in the Methods chapter, my pronunciations of the “same” sounds do not overlap to the same extent as Lisker's pronunciations, reflecting the likely fact that we did not pronounce them in the same way, and thus that these “same” sounds are not really the same at all, for the different speakers involved. Nonetheless, Lisker's point is important. Vowels which occupy the same location on a standard formant chart can sound quite different. A good example of this is the pair /i/ and /ey/. /i/ glides inward and down, and is an “intrinsically short” vowel, while /ey/ glides forward and up, and is “intrinsically long”. These vowels glide in different directions, for different durations. However, these vowels frequently glide through the same area of formant space. Thus the formant frequency measurements at the “nucleus” may be identical for the two classes despite the fact that they don't sound identical at all.

Overlap in formant space does not necessarily imply auditory identity of the overlapping tokens. On the other hand, small differences in formant space can correspond to very clear and audible phonetic differences, as shown in the study of /pæp/ and /pep/ discussed above (page 118). The classical studies of Flanagan (1955) show how small an audible difference can be: in the range of 50Hz for F2 and of 30Hz for F1.

The data presented in this thesis contains vast amounts of overlap between phonemes. Other acoustic features besides the instantaneous measurements of the first two formant frequencies may still distinguish the overlapping sounds. While this overlap will in fact
cause considerable trouble for the most simple-minded of speech processing algorithms, it need not alarm us. Overlap does not imply neutralization!

On the other hand, when considering different allophones of the same phoneme that are measured in a consistent way, one may infer, from measurable and statistically significant differences between distributions in formant space of two allophones, that there is some real phonetic difference made between them. A significant difference in formant space between linguistically characterized sets of sounds amounts to a proof that an allophonic difference exists.

We will proceed in this way, using this rule of inference to argue from the documented distributions on F1-F2 charts to inferred phonetic or phonological processes.

### 9.5 The Shape of Vowel Space

The first and second formant frequencies are measured at the phonetic nucleus of the acoustic vowel corresponding to 1890 vowel tokens taken from a conversation between Vince and Santa Ana.\(^{18}\) Figure 9.1 shows the general shape of LACE F1-F2 space.

Only a few general facts can be learned from a chart like this. First, the overall distribution is roughly triangular, though it has a more vertical back edge than, for example, the Jamaican Creole (JC) vowel space, suggesting a phonetic fronting of the higher back vowels in LACE. This confirms the impressionistic finding above that the phoneme /u:/ is realized well to the front of [u].

It is interesting to make a slightly more detailed comparison with one of the corresponding charts, made for Jamaican Creole vowels (cf. page 166). There, Juba's vowel space has a mode in the [a] region, and the distribution fans up in two directions, front-and-up and back-and-up, leaving a sparsely populated area in the high-central region. In contrast, the LACE vowel space has two modes, high-front and high-back, with a lesser density of tokens occurring in the low-central area. One might summarize this difference by saying that LACE vowel space is more top-heavy than JCE vowel space. Also, LACE has less of a trough in the density of tokens in the high-central region than JC has.

\(^{18}\)For justification of the phonetic importance of F1 and F2, see the Acoustics chapter (e.g., page 21). For the definition of acoustic vowel, see page 109. For the measurement procedure, see the Methods chapter (page 135ff).
Figure 9.1: 1890 raw measurements of F1, F2.

L.A. Chicano Vowel Space: Vince, 29
9.6 Mean nuclei in F1-F2 space

Here we examine the locations of means of the measurements of F1 and F2 for each phonological vowel class. A particular mean location can be expressed with much greater precision than is justified given the data it is based on. Thus a mean might be calculated to four significant digits, while the standard error of the sample it was taken from has a spread of two significant digits, so that the actual precision of the estimate of the mean is much weaker than the 4-digit calculation would imply. For this reason the by-now-familiar bootstrap technique discussed in the Methods chapter is employed to show how much error is inherent in estimates of F1, F2 means. For a given vowel, a new sample of data is constructed by randomly selecting with uniform probability, n elements from the actual sample of n measurements, with replacement. The mean of this new sample is calculated. This procedure is repeated by a computer hundreds of times, and these re-estimated means are plotted. The re-estimated means are themselves in a bivariate normal distribution, which is a near-optimal estimate of the sampling distribution of the original sample mean. Thus the distribution of re-estimated means itself shows the area in which the true mean of the class is likely to fall, given the set of actual measurements. The chart in Figure 9.2 displays 200 re-estimated means per vowel class.

The information presented in a chart like Figure 9.2 is extremely rich and informative, but requires considerable explication to see the patterns. The phonetic grammar of Jamaican Creole vowels (page 177) is an excellent example of how the locations of vowel nuclei on this kind of chart can be predicted from quite general principles, applied in a partly language-particular way. Such perfection in a phonetic grammar may be difficult to attain in every analysis, but the goal at each point is the same: to see what the rules are that govern the relationships between the phonological structure and the phonetic forms.

Several general rules can be proposed to explain the relations among the distributions. The fundamental assumption in interpreting these data are that vowels with the same nuclei and different glides (or no glide) are underlyingly identical, and thus that the phonetic differences between corresponding nuclei must be accounted for by phonetic implementation rules, preferably general rules that apply to natural classes. The following section discusses the phonetic realizations of each of the vowel classes shown in Figure 9.2, while
Figure 9.2: Bootstrap estimates of mean F1, F2 for each vowel class

LACE (Vince): 200 Re-estimated means for each vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 9.3: Phonetic Implementation Rules for Los Angeles Chicano English

1. The nucleus of /ay/ is fronted.
2. Short non-low vowels are lowered and centralized.
3. High, non-front vowels are fronted.
4. Nuclei before /r/ are backed except /ar, er/.
5. The low-front vowel /æ/ is raised relative to the low-back vowels.

arguing for the phonetic implementation rules presented in Figure 9.3.

The basic hypothesis of underlying identity for vowels with identical nuclei seems transparently born out by the /a-/ vowels. In the phonological structure of both LA Chicano English and also of Reference American the low-back nuclei of the vowels /a, aw, ay, ar/ are identical. For Vince, /ay/ is relatively raised and fronted, but /a, aw, ar/ form a single, undifferentiated cloud, showing that their nuclei cannot statistically be distinguished in this data.

This is consistent with a view of phonetic grammar in which the phonological classes have a target phonetic realization determined entirely by their phonological content and general principles of phonetic form, plus some language-particular rules which modify the phonetic implementation of some classes according to phonological context. If the phonological content of the nuclei of two vowel classes is the same, and there are no language-particular and phoneme-particular rules modifying their realizations, then their phonetic nuclei will also be identical. This seems to describe the nuclei of these three /a-/ vowels, /a, aw, ar/, which are realized identically in phonetic vowel space.

The front-raised character of the /ay/ class might be attributed to a phonetic rule governed by stress. /ay/ is much more frequently unstressed (61%, n=193) than /a, aw, ar/ (24%, n=104). (The chi-square significance of this difference is p<0.001.) If nuclei assimilate toward their glides when unstressed, (as suggested, for example, for Alabama /aw, ow/ on page 8.3, or by the relative fronting of the effect of stress on this very vowel, /ay/, in Vince's speech shown below in Figure 9.6, upper chart, page 9.6), then the higher frequency of unstressed and thus glide-assimilated tokens for /ay/ may account for its separation as well as the direction of its separation from the rest of the class of /a-/ vowels. However, not all of the difference between /ay/ and the other classes may be
attributed to this effect. An examination of the differences between the stressed nuclei (the tails of the arrows) in Figure 9.6 below in Section 9.7.2 shows that the stressed-only means of the /ay, a, aw, ar/ classes retain much the same relations as they have in this chart which includes both stressed and unstressed tokens: stressed /ay/ remains somewhat raised and to the front of the other (stressed) /a-/ vowels. At the same time it lies to the back of the nucleus of the low-front vowel /æ/, so that one cannot, by analysing /ay/ as low-front instead of low-back, escape positing a phonetic rule to shift the nucleus of this phoneme. In either analysis, a phonetic rule is required. The contact of Chicanos with Anglos, who are somewhat influenced by the speech of (immigrating) Southerners (as are many Californians) who have a fronted, monophthongized /ay/, might be responsible for the acquisition of a fronting rule for this phoneme, but this is mere speculation. Still, a phonetic fronting rule is necessary to reconcile the basic hypothesis that the nucleus of /ay/ is phonologically identical with the nuclei of the other /a-/ vowels. This is given above as rule 1.

The phonetic identity of the nuclei of the other /a-/ vowels need not be stated in any phonetic rule; it follows from their phonological identity, and from the lack of rules to separate them phonetically.

Next, let's examine vowel length. The relationship of long and short vowels in this dialect is qualitatively quite similar to that found in Jamaican Creole: The short vowels are lower and more central than the nuclei of their corresponding long vowels. Thus /ɪ/ is backer and lower than /iː/; /ɛ/ is backer and lower than /eː/; /u/ is fronter and lower than /uː/, and /ʌ/ is fronter and lower than /oː/. A single phonetic implementation rule captures these differences, rule 2, above.

This same rule occurs in Jamaican Creole (page 178, rule 4), but there it applies to low vowels as well. The opposite formulation, that long vowels are raised and shifted to the periphery of vowel space relative to the corresponding short vowels, is equally valid; which direction the rule goes in is not yet clear. Long-vowel raising and short-vowel lowering are two sides of the same coin, in this dialect.

A redundancy rule was proposed in Phonological Preliminaries (page 88) that adds to vowels the features [back] and/or [round]. The central and back vowels are underlyingly unspecified for the frontness feature, and this redundancy rule specifies some of them (/u;
ur, or, oy; u:, o:, o:/ as phonetically [back]. The fronting of /u:, u/ may therefore be seen as undoing the results of this rule, or as preventing its application in the first place. In this chart the back-rounding rule's effects are only seen in the realizations of /o:, or/.

The fronting itself follows Labov, et al.'s (1972) third rule of chain-shifting: Back vowels front. A synchronic analysis that minimizes the number of rule-applications required to derive the phonetic system shown here would prevent the application of the backing/rounding rule to high vowels. However, as seen in the discussion of the effects of consonants on vowels, there are back, rounded /u:, u/ tokens, so it seems preferable to make the synchronic analysis parallel to the historical analysis by stating that these back vowels are fronted in some (most) contexts. Rule 3 above reflects this conclusion.

Notice next that the nucleus of /ʌ/ is much more centralized than the corresponding non-front, mid, short vowel in Jamaican Creole (there written /o/, but consisting of the same class of words, Wells' STRUT class). While the corresponding vowel in Jamaican is truly back, on the back edge of the vowel space (see page 174), this vowel in LA Chicano English (as in many other dialects) is central: it lies directly above the /a-/ vowels and well to the front of /o:, o:/.

At the same time, the high-back vowels, /u:, u/ have fronted to some extent; their means lie directly above /ʌ/ and well to the front of /o:, or/, for example. The phonetic differences and similarities between California Anglos and Chicanos are largely unstudied, but this fronting may be related to the phonetic norms of the matrix Anglo community. Some early unpublished work of mine on the vowels of one Southern-California-raised Anglo found extreme fronting of the back vowels, so that F1-F2 measurements of nuclei of /u/ and of /ɪ/ overlapped considerably; no phonetically back mid or high vowels occurred at all except before /r, ɪ/. This evidence is only suggestive, however, and further work is called for to make clear the phonetic effects of ethnic diversity and (non-) contact in California.

There are important segmental conditioning factors in this fronting process, to be

---

19 Of the remaining back-round vowels, /ɔ:/ has merged with /ʌ/, and /ur, oy/ are too infrequent in this data to display in this form (n=0, n=5, respectively).
20 This supports the point made with respect to Reference American (page 88) that /ʌ/ is not back but central.
21 A pre-medical student at UC San Diego, she is an upwardly mobile middle-class speaker. She participates in, though probably is not among the leaders of, ongoing Anglo sound changes.
explored in future work (however, see Chapter 10 for the apparently retarding effect of following /l/).

One thing that seems clear about this fronting is that it applies to a natural class of vowels, not just to one phoneme. While many vocalic sound-shifts seem to operate on a single phoneme at a time, both high-back vowels /u:/, as well as /u/, are fronted in LACE (as compared with /u:, u/ in Jamaican or as compared with /u:, v/ in the Chicago speech of Jim C., to take two examples — see pages 206, 174), as stated in rule 3 above (page 257).

Next consider rule 4, which describes the backing of nuclei before /r/. Nuclei of Vr vowels are not always identical to the nuclei of corresponding short vowels in Figure 9.2. While /ar, er/ are indistinguishable from /a, e/, respectively, /ir, or/ are both distinct from their corresponding short vowels, and fall below and to the back of the corresponding long vowels /i:, o:/.

Similarly, /ə/ (symbolized "R" in Figure 9.2) lies to the back of /ə/. If one were to identify the phonetic target of /ir, or, ə/ in F1,F2-space as phonologically related to the phonetic target of /i:, o:, ə/, then the effect of /r/ in the glide slot of the syllable would seem to be a uniform phonetic effect of backing the nucleus. This analysis could be generalized to the relationship between /e:, er/, since /er/ also lies to the back of /e:/.

(This would suggest writing /er/ as /er/. However, the phonetic difference is greater than in the other cases, and the example of /a, ar/ shows that nuclei with and without /-r/ can be phonetically identical, while /er, e/ lie directly on top of one another. Thus the target of /er/ would seem to be identical to that of /e/, and the rule of backing before /-r/-glides is restricted to /i:, o:, ə/. The fact that /er, ar/ do not undergo the backing rule appears to be a rather unnatural (and presumably dialect-specific) condition which blocks the application of the rule, as represented in rule 4 above.

Next, consider the realization of the /ə/ phoneme. This phonologically low-front vowel is not as low as the /a-/ vowels in Figure 9.2. Phonological Preliminaries claimed that [front] is the marked value while underlyingly unmarked vowels are central or back. Only one nucleus can be the lowest in the system. One may suppose that the lowest vowel should be unmarked except for [low]. Low nuclei that are phonologically unmarked for frontness or backness are in a sense purely low and nothing more, while the marked low nuclei, though low relative to other non-low vowels, are not purely or exclusively low, since
other features are intrinsic to them. One may suppose that "pure" low vowels should be lower than low vowels that have other phonological features as well. This may explain why the nucleus of /æ/ is raised relative to the nuclei of the /a-/ vowels. These suppositions are consistent both with this datum, and with the analysis of the back and central vowels of Reference American (and by extension, of LA Chicano) as unmarked.

However, it should be noted that /æ/ in this dialect is much lower (as [æ]) than it is in many other dialects, such as Chicago White English, where it may be realized as [e3], higher than the realization of /ɛ/. So the raising of /æ/ is relatively small in this dialect.

Consider finally the location of the syllabic /l/, which overlaps with /u/, to the back of /æ/ in the central, upper-mid region of F1-F2 space, as shown in Figure 9.2. /u/ overlaps /æ/, /ə/, and /l/, which lie without mutual overlap in that order from front to back.

Since F1-F2 values directly reflect properties of the shape of the vocal tract, namely of the degree of mouth-opening and of tongue-body frontness and lip aperture, and since they very closely match fine impressionistic characterizations of phonetic vowel quality, these patterns of mean formant frequencies are not epiphenomenal. They are patterns of phonetic behavior that presumably reflect the phonetic intent of speakers. Thus the rules given discussed above may be inferred to have some psychological, as well as merely behavioral, reality; thus I refer to them as principles of phonetic grammar. The discussion above has thus given a complete (though qualitative) characterization of the phonetic grammar, through which the average nuclei of LA Chicano vowels are realized acoustically.

### 9.7 Stress Reduction

All vowels in the analysed section of speech were impressionistically coded for phrasal stress as discussed above in the Methods chapter. This section examines the role of stress in phonetic implementation by considering the effects of phrasal stress on the phonetic quality of (acoustic) vowel nuclei. In most of the stress studies in this thesis, the results presented are limited to the means of the distributions of stressed and unstressed vowel classes. Here I would like to give the reader a feeling for the degree of overlapping that is involved between stressed and unstressed subsets of a vowel class; this is not done for

---

22 Acoustic vowel is defined on page 108.
other chapters. So the first section below discusses sets of charts of raw measurements, while the second discusses the relational patterns using only the means, as is done in the other chapters.

9.7.1 One Vowel at a Time

Tables 9.4, 9.5 show the effects of stress classifications using raw formant frequency measurements. Vr vowels are not included. 0 represents unstressed tokens, while 1 represents primary stressed tokens and 2 represents secondary stressed tokens. In the chart headings, /i:, u:, e:, o:, æ, ʌ, e/, are written as /iy, uw, ey, ow, ae, /, a, E/, respectively.

The effect of stress is to lower the vowel’s realization for phonemes /æ/, /ɛ/, /e:/, /æw/, and possibly /ʌ/, /a/, and /ay/; backing applies to /o:/, /u/, and perhaps /u:/, and /oy/(and /o/ in Spanish words); fronting may apply to /i/.

The only vowel for which no effect is evident in these charts is /i:/, for which stressed tokens appear everywhere that unstressed tokens do, and in apparently similar proportions.

Syllabic /r/ (that is, /œ/) is more spread out in all directions when unstressed than when stressed; that is, there appear to be target F1-F2 values for stressed /œ/, whereas when unstressed that target is much more variably realized. With relatively few tokens, /a/ shows the opposite effect: stressed tokens are distributed more widely than unstressed tokens, which occupy only a small part of the range of the stressed tokens.

The general picture fits nicely with the standard qualitative description of the effect of stress on vowels in English: stress generally makes vowels more peripheral and helps vowels to attain their targets. Thus front vowels should be fronter, back vowels backer, and low vowels lower. Further, if stressed vowels are more likely to hit a phonetic target than unstressed vowels, then vowels should be less widely distributed when stressed than when unstressed. This is generally the picture provided by the above charts.

An exception should be noted. Stressed and unstressed /æ:/ are equally widely scattered along the F1 dimension, though stressed tokens are lower than unstressed tokens. So for this phoneme, phrasal stress does not make the vowel more likely to attain a particular target, since there is no evident target in either stressed or unstressed sets.

262
Figure 9.4: Stress effects in raw F1-F2 data, #1.

Chicano English /iy/: Stress Level

Chicano English /1/: Stress Level

Chicano English /uw/: Stress Level

Chicano English /U/: Stress Level

Chicano English /ey/: Stress Level

Chicano English /E/: Stress Level

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 9.5: Stress effects in raw F1-F2 data, #2

Chicano English /aw/: Stress Level

Chicano English /^/: Stress Level

Chicano English /ae/: Stress Level

Chicano English /a/: Stress Level

Chicano English /aw/: Stress Level

Chicano English /ay/: Stress Level

264

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
9.7.2 All Vowels at Once.

We can reduce the data discussed above to a single chart if we display only the means of the stressed and unstressed distributions for each vowel. Obviously, reducing clouds of data to a few points will radically change the picture: the overlap of the distributions will not be shown at all, though as we have seen, that overlap is considerable. At this cost, we are able to summarize the effects of stress on all the vowels.

Notice that so far this presentation of the effects of stress has included clitic words with non-clitic words. Many clitic words are likely to have a phonological schwa in place of the phoneme which occurs in stressed isolated pronunciations. That is, "that", when pronounced [ðæt] should not be classed phonologically with "that" when pronounced [ðət]. Because of possible confusion over what phonemes occur in many clitics items, two charts are shown in Figure 9.6: with vowels in non-clitic words only, and with vowels in both clitic and non-clitic words.

The means of stressed (primary or secondary) and unstressed allophones of each vowel class are plotted with an arrow from the stressed mean to the unstressed mean. Dashed arrows signify an insignificant difference while solid arrows are statistically significant (5%, two-tailed t-test, cf. page 5.9.2). Vowels lacking multiple tokens of both stressed and unstressed allophones in non-clitic word English words are displayed at the mean of the class that does occur, with an apostrophe, ' , signifying stress (as with /aw , er/), and no mark signifying unstressed (as with "L", which stands for syllabic /l/).

I will discuss just the upper, clitics-excluded chart, because it exhibits the clearest pattern. The means of the distributions of all the non-low vowels, /iː , ɪ , ʌ , ϕ , ʊ , ə , oʊ , oy/ , shift quite precisely in the direction of a single focal point. Although some of these individual effects are insignificant — at a rather stringent level — the fact that all of the arrows follow a consistent pattern suggests that a single, quite significant process applies to all of them. The arrows point fairly accurately in the direction of a single location in F1-F2 space; this suggests the metaphor of a reduction "target". This target in LACE

---

23For the precise definition of this distinction see page 5.5.
24It should also be noted that the word you, which occurs quite frequently, was phonologized as /yə/. It never sounds back or rounded, even when stressed. The single token of stressed /ə/ that occurs in the clitics-included chart is from the word, "you".
25Two-tailed t-test, 5% level of confidence, as described on page 153.
Figure 9.6: Effects of impressionistic stress in Los Angeles Chicano English. Arrows connect means of stressed (tail) and unstressed (head) tokens.

LACE (Vince): Effects of Destressing.
Clitics excluded

LACE (Vince): Effects of Destressing
Clitics included

Dashed arrows are insignificant on 5%, 2-tail t-test

266

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
corresponds to a vowel quality close to [i]. For these vowels, stress reduction should be described as a single process which applies to all of them in the same way, by shifting them (on average) in the direction of a maximally neutralized vowel, which corresponds closely to the F1-F2 location of [i].

Exceptions to this aggregate process of shift toward the target [i] are two: the class of low vowels, and /e:, ir/. /e:, ir/ may be ignored since the effect of stress on them is insignificant. /e:/, for example, is quite widely spread out in the F2 dimension, with stressed allophones as a whole forming a row just below the unstressed allophones. The great variance in the F2 dimension, plus the least amount of skew in the data introduced by other unbalanced factors could account for the fact that /e:/ does not shift in the expected direction.

The reduction process does not apply with such clarity to the class of low vowels shown in the upper chart. This may be due to the paucity of non-clitic data for these vowels (5 unstressed /ar/ tokens, 3 each of unstressed /a, ar/, none of /aw/). They do not show the same degree of inconsistency with the general process in the lower chart, where clitics are included, and these vowels all raise directly upward.

The third exceptional low vowel in the clitics-excluded chart, /ay/, has a more traditional characterization: the nucleus is shifting in the direction of the glide. This amounts to increased coarticulation between the nucleus and glide in the unstressed forms of this vowel. (Similar effects also occur with certain gliding vowels in Alabama and Chicago, pages 241, 213).

Does reduction shrink vowel space, or does stress expand it? Some vowel nuclei overlap much more when unstressed than when stressed. A basic structuralist principle says that if two classes merge in some particular context, then the distinction between the two must be present underlingingly, and neutralized in that context, rather than being absent underlingingly, and produced unpredictably in the contexts where it does appear. The underlying, or “true”, form is the distinct form, while the derived form is the merged one. This principle can be used with /u/ and /uː/, for example, to say that the stressed forms are the “true” or underlying targets, while the unstressed forms which appear to be nearly neutralized (in this representation), are derivative. Similarly, all vowels are relatively neutralized if they all shift towards a common reduction target, and thus towards each
other. If this interpretation is followed, the correct characterization of phonetic vowel reduction is not the vowel space expands when stress is applied, but that vowels move in the direction of [r] when stress is reduced. Thus the following rule of phonetic grammar.

When phrasal stress is reduced, the phonetic realizations of vowels shift towards the reduction target.

The grammatical characterization of this process is a subtle matter. The location of the phonetic realization of /a/ is especially significant here. The reduction target may be identical to the phonetic realization of /a/; certainly it is seductive to simply characterize vowel reduction as shift towards /a/, though there are hidden complications in such a statement. The shift is actually towards the phonetic realization of /a/, not towards [a], which is mid-central, not high-front, like this target, and not towards /a/, which is a phonological entity (written with / /'s) rather than a phonetic one. Statistical tests that may require even larger quantities of data would be necessary to distinguish the realization of /a/ from the global target of vowel reduction. Such a test would have as a null hypothesis that the two are identical. Until such tests are made, I will assume that the realization of /a/ is in fact identical to the reduction target.

Next, consider the phonological treatment of /a/. It is analysed as underlingly absent, or featureless in Phonological Preliminaries. That is, it is unspecified for height, backness, or stress, and may even lack an underlying Nucleus position. Essentially, this analysis states that /a/ is a phonological vowel that is nothing at all (that is, it is constituted by no underlying features, only relatively indistinct phonetic features). So /a/ doesn’t actually exist, only its realization does, which is entirely a creation of low-level processes of phonetic implementation, such as processes that insert a Nucleus slot where none is present, in order to satisfy syllable-structure well-formedness conditions.

If the Nucleus slot is inserted in an unstressed syllable, and is filled by no other features\textsuperscript{26}, then without further specification, the vowel will be realized with the phonetic quality of the reduction target, which is the only source (besides segments in the adjacent environment) of phonetic vowel quality for this vowel segment. Finally, the ancient, but

\textsuperscript{26}For example, /\alpha/ is underlingly a Glide slot specified with the feature [rhotic]. This feature is then linked, after nucleus-insertion, to the nucleus timing slot. See Phonological Preliminaries, page 85.
still hardly confirmed idea of Sievers (1876), that different languages have different rest positions for the vocal tract may be relevant here. The reduction target could be not just the realization of the unspecified vowel, it could also be the rest target. It appears that this target is different from those for other dialects, as summarized in the concluding chapter.

This section has shown a clear pattern of vowel shifting that correlates with phrasally (or lexically) destressing a vowel. The pattern seems to be best explained as a phonetic rule of vowel reduction which shifts nuclei in the direction of the "reduction target", a location in F1-F2 space that corresponds to impressionistic [t], and which we may assume is identical to the target, or mean location, of the phonologically reduced vowel /ɔ/.

9.8 The Low-Back Merger

The merger of /a/ and /ɔ/ is reported to be "'the norm in Canada, and is spreading with great rapidity in most areas of the USA', despite being 'unknown in England'."27 This merger may be found in Chicano English, since it is claimed28 to occur in the Anglo speech of California. However, "the low back merger has been the subject of surprisingly little systematic study" (Herold, 1990:8). Most previous work in the US amounts to repeated anecdotal observations and claims that the merger has occurred, while some studies purporting to bear on the issue of the merger (Johnson 1974, Moonwomon 1987) examine only the /ɔ/ class, and thus can say nothing about its patterning relative to the /a/ class. Thus a systematic comparison of the acoustic patterning of the /a/ and /ɔ/ classes in another California dialect would be a positive contribution to the largely anecdotal literature on this merger.

Among the vowels in the conversation studied are 53 tokens of the /a/ class, and 28 tokens of the /ɔ/ class. Measurements of the formant frequencies at chosen nucleus locations are presented in Figure 9.7. Some preliminary discussion is necessary to interpret this chart.

Herold (1990) notes that the phonological environment of these vowels is highly skewed: /ɔ/ occurs much more often than /a/ before /l/ while /a/ occurs much more often than

---

28cf. DeCamp, 1959; Johnson, 1974; limitations of these studies are pointed out in Herold (1990:8-11) and are overcome in systematic instrumental work by Moonwomon, 1991.
/ɔ/ before /p/. Since /l/ has a strong backing effect on preceding vowels, (see below: Effects of Adjacent Consonants), and the phonetic realization of /ɔ/ is generally phonetically backer than that of /a/, confusion is possible between the effect of vowel identity and the effect of the phonological environment. Thus we also display the following consonant along with the tokens of the /ɔ/ class (Wells' lexical sets, THOUGHT and CLOTH). Since /ɔ/ tokens (Wells’ LOT set) are much more common, and almost never precede /l/, we display them here with just a “,” and without the following consonant.

The measurements of /ɔ/ tokens are displayed with the symbol “c”, plus the following consonant. As may be seen, /ɔ/ tokens are generally higher and backer than /a/ tokens. This difference is consistent with the view that the merger has not taken place: the difference could be due to the underlying phonological distinction between the two lexical classes.

However, notice that following consonants for the raised and backed tokens are few: /l, n, η/. Since /l/ is known to have a backing-and-raising effect on preceding vowel nuclei (again, see below in Effects of Adjacent Consonants), since nasals are known to frequently have a raising effect on preceding vowels,29, and since the raised tokens before /n/ are in the clitic word, on, it remains possible that these differences may be attributable to other allophonic or coarticulatory effects. Since the remaining THOUGHT-class tokens, with following /s, z, t, j/ occur well within the main body of the LOT distribution, it is likely that the difference of distributions of the two lexical classes is not due to an underlying lexical distinction.

It would appear that more data from more speakers, targeted towards answering this particular question would be useful in making certain whether or not /ɔ/ and /a/ are merged in Chicano English. However, the evidence here, though superficially supporting the hypothesis that they two classes are separate for this speaker, on a closer examination do not support that hypothesis. It would appear that this is a Low-Back Merged dialect of English.30

29See, for example, Labov, Yaege, & Stein (1972:94).
Figure 9.7: LOT vs. THOUGHT, with following consonants.

Chicano English: LOT (.) and THOUGHT (c) Lexical Sets

The following consonant is also plotted.
9.9 Summary

This chapter has proposed a structural analysis of the surface vowel system of Los Angeles Chicano English. The average phonetic realizations (in F1-F2 space) of the acoustic nuclei of the phonological vowel classes were described by several rules of phonetic implementation. Then the effects of phrasal stress on the performance of vowel production by one speaker were investigated through F1 and F2 measurements, finding that stress reduction shifts vowels in the direction of a high and somewhat front of center reduction target, which could be identical with the phonetic realization of /ə/. Some dialect-specific effects of following consonants are also discussed in Chapter 10.
Chapter 10

Conclusion: Consonant Effects, and Phonetic Grammar

This concluding chapter returns to focus on theoretical questions about phonetic grammar, a controversial object of description. It is a fact that physically motivated phonetic processes are usually not physically necessary, and can vary stylistically, so that the most natural, physically easiest and most simplified phonetic forms are restricted to certain styles. There are similar differences in these details across dialects. "Hard" coarticulation does exist, since the tongue cannot move infinitely fast. However, natural phonetic processes, while physically simpler or easier, are not physically necessary.

The tongue is nearly all muscle, and the only mass it must move is itself. Further, a considerable part of the human brain is devoted to the physical control of the tongue. So it should not be surprising that the coordinated movements of the human tongue are among the most complex and rapid large-scale physical actions naturally performed in the animal kingdom.¹ The physical limitations of the human apparatus of speech — the finite mass and amount of force that may be exerted on that mass — are real, but there is considerable leeway for linguistic control (and conscious control) of details of articulation at normal speaking rates. The phonetic effects which lie within this realm of "soft" coarticulation

¹The wing movements of hummingbirds and bees, for example, though more rapid, appear to be less complex (as well as smaller in physical scale), since they follow a single, repetitive pattern of motion, while the human tongue moves in leaps and contortions that, while not entirely unpredictable, are not yet well understood.
are not due to absolute physical constraints. People learn coarticulations that are often physically easy but not always physically necessary. As argued in this thesis, the control of these often intricate details is part of the human language faculty, because it is part of what differentiates one dialect or language from another.

The system by which surface phonological structure is interpreted phonetically is a partly linguistic system (Liberman & Pierrehumbert 1984). This system includes the learned rules of soft coarticulation and of prosodically governed (stress) reduction, as well as the interpretation of abstract tonal specifications as F0 contours. One of the primary elements of this system is a specification of the target, or average, phonetic vowel qualities; these qualities do not follow from phonological structure and universal principles of phonetic implementation alone, unless the many effects described in the preceding dialect studies can be accounted for by some aspect of phonological structure heretofore unknown.

10.1 Consonant Effects

In this chapter, coarticulatory effects are investigated in different dialects, in order to show that vowels in the environment of particular following consonants are regularly found to have modified phonetic quality (relative to their average quality, in terms of F1, F2 measurements), in ways that are particular to individual dialects. The exploration of the form and structure of the phonetic system by which consonant environments exert their varied effects on vowels in the dialects studied here is beyond the scope of the present work. Instead, I wish simply to document the existence of this realm of linguistic variation. Once the existence of these dialect- and language-particular coarticulatory effects is accepted, further research may begin to characterize the structure and functioning of this aspect of the linguistic system.

In view of the distinction between “hard” and “soft” coarticulation advanced here, the system by which phonological representations are converted into speech would appear to have considerable leeway for quite complex and as yet badly-understood sets of interactions between vowels and their phonetic or phonological context. These interactions may or may not be dialect-specific, but it is important not to approach their study with the precondition that they must be universal effects. Before deciding whether a given effect
is language-specific or not, we must see what some of these effects are in the first place. To this end, a study of the effects of two particular consonant environments on the F1-F2 measurements of the vowel nuclei has been carried out. These are the backing effect of following /l/ in all dialects except Jamaican, and the lowering effect of following /η/ in Alabama as opposed to other dialects.

The statistical methods are discussed above in Section 5.9.2. To briefly summarize the discussion there, the effect of a given consonant on a particular vowel is measured by locating the means of two sets of F1-F2 measurements: first, those of the given vowel in the context of that particular consonant, and second, the measurements of that vowel in all other contexts. The data points are projected onto the line between the means in F1-F2 space, and statistical tests for different variances (F ratio) and different means (equal-variance t-test or unequal-variance t-test) between the two sets of measurements are carried out.

It should be pointed out that the formant trajectories for a given vowel would show even more differences across consonant contexts than are shown by the differences among nucleus measurements. Thus the differences found here are undoubtedly the tip of the iceberg, and a more complete theory of phonetic performance must account for those differences as well.

The task of interpreting these statistical effects is, in general, a difficult one. The fact that the data is unstructured natural speech means that there is no control over other interacting effects. For example, after a /w/-glide, unstressed vowels are frequently backed and rounded, so that the word *with*, for example, will have a nucleus far to the back of tokens of the /l/ vowel in other contexts. Due to imbalances in the data, such an effect may masquerade as an entirely different one. Thus for example if the word *with* occurs very frequently, and the following consonant /θ/ otherwise co-occurs rarely with preceding /l/, then the strong backing effect of the preceding /w/-glide will also show up as an apparent strong backing effect of /θ/. Thus it is a difficult matter to be certain, in any one case, whether the effect is genuine or due to other factors which are skewing the data. For this reason, the effects presented below are not based on single vowel-consonant coarticulatory or allophonic effects (here called VC effects), but are based on sets of individual VC effects which pattern similarly. If all the labial consonants (/p,b,m/), for example, have a
similar low-fronting effect on a preceding vowel, it is unlikely that the low-fronting effect is due to some lexical or prosodic interaction, because an identical interaction must have coincidentally applied to all three consonants. That seems to be a rather unlikely event. These two effects, chosen from among many, are consistent in this way.

The simplifying assumption of a universal phonetic implementation system may be necessary in the absence of a necessary body of observation and theory. This assumed universal phonetic system contains solely information about such matters as acoustics, aerodynamics, vocal-tract anatomy, and physiology, and additionally a non-cognitive, non-linguistic, universal system by which phonological categories are implemented and interpreted in speech production and perception. Some previous work testing these assumptions has been done. Göstë Bruce, for example, showed in his work on Swedish tone that the implementation of accent places a pitch inflection at a point in the syllable different from the implementation of stress in English. Similarly, Laniran (1991) showed that phonologically similar tone sequences on phonologically similar segmental sequences was realized in different patterns of pitch-contours in different two different languages. In these cases the phonetic realization of similar phonological forms is different according to linguistic or other high-level factors.

I will conclude that strictly phonetic differences between dialects (and languages, for that matter) do exist. Phonetics may consider cross-linguistic and cross-dialectal differences, just theoretical linguistics includes strictly phonetic considerations.

10.1.1 Effect of following /l/

The effects of following /l/ on vowels is perhaps the strongest and most regular VC coarticulation effect found in this data. However, the effects differ across dialects. Consider the plots in Figure 10.1, which display for six speakers the statistically significant effects of following /l/ on vowel nuclei.

The arrows displayed represent statistically significant effects. The head of the arrow is located at the mean of the distribution of instances of that particular vowel which precede an /l/, while the tail of the arrow is at the mean of the distribution of all other instances of that vowel.

Arrows with large arrowheads represent differences that are significant at the p<0.001
Figure 10.1: Effects of following /l/ on vowel nuclei in four dialects: JC (Jamaican Creole), AE (Alabama English), LACE (Los Angeles Chicano English), and CWE (Chicago White English).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
level. Medium sized arrowheads represent a significance of p<0.01, while small arrowheads represent those effects with a two-tailed significance of p<0.1.

The basic dark-l coarticulation pattern is that seen most clearly in the CWE chart of Rita. The database of measurements for her speech was much larger (4470 tokens) than those of the other speakers, and thus the significant effects are generally more numerous and more significant. /l/ has a strong backing effect on all vowels, the strongest effects are on the high-front vowels, /iy, ey, i/, where F2 is lowered by as much as 1000Hz. These differences are both statistically significant, and above the difference limen for formant frequency perception, so these are real phonetic effects, by the rule of inference argued for in Chapter 4.

This basic pattern is repeated in the speech of all the mainland American speakers, applying to all the significant effects for Judy, James H., and Vince (as well as Jim, not shown), with only a few exceptions in which backing co-occurs with rather extreme raising or lowering. Some of these exceptions may be attributed to other processes, either phonological or phonetic, while some of the them appear to be unexplained except as brute idiosyncrasies of the phonetic implementation system of the individual dialect. Each of these exceptions is discussed below.

The first exception involves the /el/ sequence in Chicano English, which is shown to have a slight backing effect, but primarily a very strong lowering effect on /e/. /l/’s effect on a preceding /e/ is a marker of Chicano speech in the Southern California. Among the words found with this alternation are [ẽ]lementary, B[ẽ]lvedere, hims[ẽ]lf, r[ẽ]latives, w[ẽ]ll, t[ẽ]ll, etc. This phenomenon was first noticed by Metcalf (1979), and has also been discussed by Wald (1984), and Santa Ana (1991). The linguistic interpretation of this significant shift (p<0.0048) is unclear, but it is possible to attribute it to a change of phonological category (cf. Santa Ana 1991). It is clear that the effect of /l/ in the development of the ethnic Chicano dialect was quite different when it occurred in the context of /e/ than elsewhere, but now this effect may be phonologized. If the vowel in this environment is taken as underlyingly /e/, then a phonetic rule is required to lower the vowel in this context to [ẽ]. But if it is /ẽ/, then no such rule is necessary.

Another apparent exception can be understood via the interaction of several phonetic processes. Jackie from Chicago has a sharply lowering effect of /l/ on /ay/, as seen in
her chart in Figure 10.1. This effect is significant at p<0.035, even though it results from only two tokens, *Niles* and *I'll*, which are at the low back periphery of the /ay/ distribution. These tokens are in this location because of three phonetic effects: relative lowering due to stress, relative lowering due to the following voiced environment, and backing due to following /l/. Chicago (and Northeastern U.S. speech more generally) displays an alternation in which the nucleus of /ay/ is raised before voiceless consonants. Following /l/ is a non-raising — i.e., lowering — environment for this reason. Further, stress reduction has a raising effect on /ay/ as seen in Figure 7.10, so that stressed tokens are relatively lowered. Finally, the backing effect of /l/ does seem to apply to these tokens in addition to the other lowering effects.

The chart of the Alabama English speaker, James II., in Figure 10.1 contains several phonetic anomalies. For /e/ and /ow/, in addition to a backing effect of /l/, there are also a strong lowering and raising effects, respectively. The lowering of the nucleus of /e/ before /l/ occurs mostly with preceding labial environments: *well, twelve, fellow*. The effect may be due either to the following /l/ or the preceding /w/, but in either case it is an unnatural (non-assimilatory) phonetic effect. Coarticulation with a high-back-round glide, /w-/ , should have a raising and backing effect, but this effect is of lowering. Similarly the strong low-backing effect (p<0.0001) cannot be attributed to the general effect of /l/ as found in other environments or in other dialects. The other case is the strong back-raising effect of /l/ on the nucleus of /ow/ in James II.'s speech (p<0.0007). This also cannot be attributed to stress, which is relatively balanced (9/14 tokens are stressed), or to preceding segments (either consonants or vowels) which fall into no natural class, even in a majority of cases. The words involved are *old, told, hold*, and *hole*, in order of frequency.

Many vowel-consonant interactions in this and in other data not presented here are phonetically unnatural patterns which do not appear to be explicable in terms of more general processes. Until a better explanation can be found, the only way to represent them appears to be as an arbitrary, language-particular stipulation in the system of phonetic implementation, such that these particular contexts have these particular arbitrary phonetic effects. Examples of this kind could be multiplied ad infinitum. It may be hoped that further exploration of this class of facts may lead to deeper understanding of the interaction of various processes of phonetic implementation.
A deeper pattern is evident in the effects of /l/ on the vowels in Jamaican Creole. In contrast to the mainland speakers, the Jamaican Creole speakers have a qualitatively different pattern of effects. In their charts, shown in Figure 10.1, the usual backing effect of /l/ appears to apply to the back vowels, but the front vowels do not show this effect, and have raising instead, where the more significant effect, in Juba's chart, includes some fronting as well. The long low vowel /aa/ for Roasta has a quite significant, though small, low-fronting effect, opposite to the backing effect in other dialects. This fronting effect may derive from an interaction: most tokens are of the words all, always, and 29/34 are stressed; stress has a small fronting effect on this vowel for Juba, though not for Roasta (cf., Figures 6.9, 6.10).

The evident explanation for the lack of a backing effect on the low and front vowels is the quality of the /l/ itself. Jamaican /l/ is quite different from the /l/ found in other dialects, reflecting the differences in phonetic realization of this phonological category. In Jamaican, even tautosyllabic /l/ is clear, non-velarized, while in the other dialects studied in this thesis, tautosyllabic /l/ is dark, velarized, often vocalized. Because the realization of tautosyllabic /l/ in mainland dialects includes a velar constriction involving the body of the tongue, it has a strong effect on preceding vowels, which also derive their quality from movements of the tongue-body. Where /l/ is not velarized, there is no such effect, as with the high-front and the low vowels in JC. /l/ appears to be redundantly velarized in the context of high-back vowels, which involve a high-back (that is, velarized) tongue position to begin with. Velarization would seem to be a redundant phonetic feature added to the form of the realization of /l/ in JC in the context of high-back vowels, which in turn has a high-backing effect on the vowel nuclei.

How are these effects to be accounted for in linguistic theory? /l/ is phonologically a lateral continuant in English dialects. The /l/’s of Jamaican, of Chicago White English, of Los Angeles Chicano English, and of Alabama Southern White English are all lateral continuants. An underspecified representation for /l/ would contain solely the feature [lateral], which distinguishes it from /r/; other features, [continuant], [voice], etc., are predictable and redundant. Despite these underlying phonological similarities between the /l/ of mainland American dialects and the /l/ of Jamaican, their phonetic realizations are quite different. Are their phonological forms also different? One may imagine that the /l/
of the mainland dialects has a redundant phonological feature, [velar] or the like, while Jamaican Creole has no such feature. However, Fujimura and Sproat (1989) demonstrated that clear-1 and dark-1 in English (in some mainland U.S. dialect) is a gradient phonetic difference that depends probabilistically on syllable structure. The clear-1, dark-1 distinction in this dialect is a probabilistic, gradient, therefore phonetic effect, rather than a categorical, discrete, phonological alternation. Therefore there is no need to include a categorical phonological feature like [velar] or [dark], etc., in the surface phonological representation of mainland /l/, since /l/’s degree of “darkness” is necessarily determined by factors in the phonetic implementation system.

This gradient phonetic effect does not appear in Jamaican Creole. Instead, another effect appears, where /l/-darkening (which I assume is measured by its backing effect on preceding vowels) is determined by the quality of the preceding vowel. The phonetic interpretation of /l/ in Jamaican as light or dark in the environment of high-back vowels may again be either a phonetic or a phonological process. But in either case, the phonetic process which, according to Fujimura and Sproat, implements /l/ in mainland dialects as more-or-less dark depending on syllable structure, is absent in Jamaican Creole. It follows, then, that phonetic implementation processes exist in some dialects that are absent in others. Furthermore, if the dark-1 which occurs after high-back vowels in Jamaican Creole is also conditioned by a strictly phonetic rule, as is plausible since it is so conditioned in mainland English, then it would seem that there are two distinct phonetic processes by which lateral continuants are implemented as sounds.

10.1.2 Alabama Lowering Before /ŋ/.

Among the clearest of “soft coarticulation” (linguistic-phonetic) effects between vowels and following consonants is that between /i/ and following /ŋ/.

The effect of following /ŋ/ is phonetically quite strong in certain cases, and in combination with certain vowels. This is reasonable in a number of ways. First, the body of the tongue must raise in order to produce a velar closure for /ŋ/; this movement is relatively slow, compared to the movements of the tongue tip, and this makes it more likely that the /ŋ/ gesture will be coarticulated with surrounding segments. Partial nasalization of a
vowel in the context of a following nasal consonant occurs frequently in English; nasalization has an obscuring effect on the formant frequencies of vowels. Finally, /ŋ/ represents a lengthening environment in English, since vowels before voiced segments are generally phonetically longer than in other environments. In an environment where lengthening has applied, there is extra time for gestures to reach their targets (relative to a following non-lengthened environment, such as before /k/), so that the restrictions of hard coarticulation are weakened, and there is an opportunity for variation in the temporal form of the gesture. Thus this environment seems a priori to be one that may well be exploited to produce stylistic or dialect-specific phonetic effects that are audible but not lexically distinctive. Consider, then, the effects of following /ŋ/ in different dialects, as shown in Figure 10.2.

Insofar as this data is capable of displaying, dialects are perfectly self-consistent, but quite different from one another. I will concentrate on the effects of /ŋ/ on preceding /i/. The opportunity that occurs in the environment of /ŋ/, to have dialect-particular patterns of soft coarticulation, appears to be exploited in the Alabama dialect, as opposed to the other dialects. In Chicago and elsewhere, the effect on the short high-front vowel /i/ (written /i/ in Jamaican) is uniformly one of fronting, and usually also raising (except in L.A. Chicano). This fronting effect cannot be due to a phonological alternation between /i:n/ or /i:n/ and /ŋ/ in the form of the present participial suffix -ing (the transcription distinguished -in' and -ing, so cases with following /n/ are not included here). Thus there is a general high-fronting effect on /i/ of following /ŋ/, which may be understood as a natural phonetic effect, since front-articulated velars might reasonably have a fronting effect on preceding vowels. (This is a complex point, however, since velars get their front-back articulation from the adjacent vowels to begin with. For this effect to actually work this way, there must be two steps, in which the consonant gets its relatively front place of articulation within the range of "velar" places of articulation from the front-back features of the vowel, and then the vowel is itself fronted by the fronted velar consonant.)

However, in the Alabama pattern of James II., in Figure 10.2, the effect on /i/ is entirely different. Here, instead of the apparently natural high-fronting effect found in other dialects, including Jamaican as well as LA Chicago and Chicago White English, the effect is an opposite, unnatural, anticoarticulatory effect. Instead of raising /i/, following
Figure 10.2: Effects of following /ng/ on vowel nuclei in four dialects: JC (Jamaican Creole), AE (Alabama English), LACE (Los Angeles Chicano English), and CWE (Chicago White English).
/ŋ/ lowers it. Impressionistically, in this speech I find that /ŋ/ sounds like [æŋ]. This occurs not just in the suffix /-ŋ/, but also in spring, finger, thing, etc.

The effect of /ŋ/ in Alabama is perfectly general within this idiolect, insofar as this data shows it: vowels preceding /ŋ/ undergo lowering. Thus /ʌŋ/ is realized as [oʊŋ], /ɔwŋ/ is realized as [ɔʰŋ] (quite different from [oʰ] which is typical in other environments). F1 rises in frequency for /ʌ/ by 118Hz (p<0.0002), for /i/ by 108Hz (p<0.00001) and for /ɔw/ by 54Hz (p<0.005). This consistent and significant pattern suggests a phonetic lowering rule for the nuclei of vowels in the context of following /ŋ/ in James H.'s speech. The pattern is consistent with the generally greater amount of gliding and diphthongization occurring in Southern speech (cf. Feagin 1991).

This is further clear evidence for dialect differences in the system of phonetic implementation. I know of no proposals that /ŋ/ itself might different across dialects, so I assume that the consonant which triggers the effect is the same as in dialects where it does not have this effect. The effect cannot instead be attributed to the phonological features of any particular subset of the vowels, since all three of these vowels show the same effect. It must be a phonetic process, characteristic of this speaker and, one may assume, of other speakers of his dialect.

10.2 Phonetic Grammar

This thesis has partly characterized the phonetic interpretation system of language through an investigation of the surface distributional patterns of acoustic phonetic measurements, correlating those patterns with features of the phonological structure: vowel classes, stress, and following consonants. It was found at all three levels that important phonetic differences exist between dialects, in addition to and separate from the phonological differences that distinguish them. At two levels, in the patterns of distribution of the mean nuclei of each vowel class, as well as in the patterns of vowel reduction, rule-systems which generate the observed distributions of acoustical measurements were characterized. Because languages differ in these phonetic patterns, and because they can be described by precise and language-specific rule systems, I believe it is appropriate to call this level of linguistic description, namely, the system of phonetic implementation of surface phonological structure
as sound, "phonetic grammar".

In this chapter, the existence of dialect-specific phonetic interpretation processes of coarticulation was demonstrated. Linguistic description and theory can now begin to explore the internal structure and functioning of this intricate system of interacting phonetic processes.

An important conclusion of this thesis is that historical phonetic rules, which restrict the range of possible chain shifts, appear to play a role in the phonetic interpretation system. Universal principles of phonetic implementation are supplemented by dialect-particular applications of these general phonetic rules, in order to describe characterize the system of average (one might say, "target") vowel nuclei. This general idea is quite similar to the basic theoretical approach taken in Liberman & Pierrhumbert (1984), an approach that was proposed in order to account for observed patterns in pitch contours. The idea of a phonetic grammar consisting of interacting universal and dialect-specific factors was taken to a logical extreme in the phonetic grammar of mean vowel nuclei was presented for Jamaican Creole nuclei in Chapter 6. Partial characterizations of conceptually similar phonetic grammars are also contained in the descriptions of the other dialects.

In the studies of vowel reduction presented in the discussions of the four dialects, there appeared to be a phonetic "reduction target", towards which vowels shift when phrasal stress is reduced. This reduction target seems to be somewhat different in different dialects. In Jamaican, the short vowels reduce in the direction of a mid-central position in vowel space. In Chicago, a high-central position is the reduction target. In Alabama (in the speaker studied), a high and relatively front position appears to represent the reduction target. The L.A. Chicano pattern seems to have a target that lies in an intermediate between the targets in Chicago White English and Alabama English: it is somewhat to the front of a central position, but not as far to the front as is the reduction target in the Alabama speech studied. It was found that the pattern of vowel reduction was fairly consistent across speakers within a dialect, so that Judy and Jim from Chicago, for example, have very similar reduction patterns, and Rita is not very different. However the differences across dialects are quite striking. Also, the application of the process seems to be conditioned by phonological vowel length in Jamaican.

We may conclude that vowel reduction as a function of phrasal stress is a widespread
phenomenon in English, if not also throughout the world's languages. However, the process is not identical in different dialects. The main difference between the dialects appears to be in the location of the reduction target, which appears to range from mid-central to high-central to high-front. The reduction target is never peripheral in the front-back dimension of vowel space, though it may be quite high. The patterns observed are consistent with the hypothesis that this reduction target is identical with the phonetic realization of /ɔ/, which may additionally represent the "basis of articulation" of Sievers (1901), which was taken up as the "neutral position" in SPE (Chomsky and Halle 1968). In this thesis, the phonological status of this position is not understood, as in SPE, as the basis for deriving the marked and unmarked values of the phonological features, though proposals might be made in that direction in the future. However, the location of this position is a crucial element in the explanation of the surface patterns of vowel reduction. Phonetic grammar must specify this parameter in some way.

The system by which surface phonological forms are implemented in phonetic production has been partly characterized for a number of dialects in this thesis. Both endpoints of the mapping performed by this system were explored in detail: in Chapter 2, the acoustic dimensions were related understandably to the articulatory configurations that give rise to them, while in Chapter 3, the surface phonological structure of the vowel system of a useful fictional dialect, "Reference American" was characterized using autosegmental structure, underspecification theory, and privative features. In the large-scale acoustical studies of surface patterns of phonetic performance conducted in each of the dialect chapters, patterns in the system of phonetic interpretation were explored and characterized, leading to considerable progress in some aspects. In addition to the characterization of abstract phonological structure and the characterization of the physical aspects of the mapping from surface phonological structure to formant patterns, various parts of the intermediate system between physical and phonological structure were explored. The existence of differences across dialects in this system was demonstrated. The examination of surface distributional patterns of acoustic phonetic measurements and the correlation of those patterns with features of the phonological structure, including the underlying vowel classes, phrasal stress, and in this chapter, certain following consonants, has led in the end to insightful, interesting, and surface-true analyses of important aspects of the system of
phonetic interpretation, or phonetic grammar.
Appendix A

Features of Time and Space in Linguistic Structure

We often speak of linguistic forms as ordered from left to right rather than from earlier to later, as though linguistic units exist only as displayed on a page and not as if they precede and follow each other in time as they are produced, understood, or structured.

An abstract feature of time, namely linearity, has always been implicitly necessary in linguistic structures. A number of abstract features of space-time, here collectively referred to as multi-linearity, are also fundamental to linguistic representations at the level of (so-called) "non-linear" phonology. Only partly formalized in non-linear phonological theory, these abstract features of space and time are here made explicit with reference to linguistic structure. Additionally, three fundamental and otherwise unrelated principles of non-linear phonological theory (the Well-Formedness Condition(s)) are proven to follow from them.

A.1 Time

Important characteristics of Newtonian time include the seven properties of immediacy, universality, continuity, staticity, unidirectionality, infinity, and linearity.

Time has a special and mysterious quality about it having to do with it occurring now, which we may name immediacy. Of an infinity of times, only the immediate instant is now. But the referent of now is constantly changing, so that it points to some different
time, for example, now. What is constant about this ever-changing temporal pointer? Not the things that occur in time, which includes just about everything. Depending on our point of view we may call these things “states”, or “events”, or “elements” or even “features” or “feature-values”.

In considering the mystery of immediacy, one may notice that its special, elusive character is not so special from an extra-temporal perspective (as in the later examination of a record of any particular temporal sequence). No event in the record is singled out as in any way different from any other event in the record, except that each occurs at its own unique time. The only remnant of now-ness left in the record is the property of each event that it occurs at its own individual immediate time, which is now relative to the event itself. This could be taken as evidence that records of temporal sequences have not recorded the essence of this principle of immediacy. However, for the purposes of examining records of events, immediacy is the principle that events occur “at” particular times; that is, that times are associated with events.

The other properties require less discussion. Universality is the principle that everything (in the universe under discussion) occurs at some time. Continuity in real time may be axiomatized as the statement that between every two (distinct) times there is another time. Staticity is the principle that an event (in this context, more properly called a “state”) may extend across (be associated with) a contiguous sequence of times. Infinity is the principle that time does not begin or end.¹ Unidirectionality specifies not just that the two directions are distinct from each other (this follows from linearity) but that one direction is special, in that certain sequences of events may not be reversed; they cannot in principle occur in the reverse order.

Linearity is defined in any mathematics-for-linguists text. The important properties here are that for a given element e in a sequence,

\[ A: \text{all other distinct elements either precede or follow } e. \]

¹Axiomatized: For every time \( t_1 \), there is a time \( t_2 \) after \( t_1 \), and a time \( t_3 \) before \( t_1 \).

²Let \( F(e) \) be a function from elements to sets of elements such that for any pair of elements, \( e_1 \) and \( e_2 \), if \( e_1 \) is in \( F(e_2) \), then \( F(e_1) \) is a proper subset of \( F(e_2) \). If such a function exists, the set is called a linear sequence. Note that \( F(e_1) \) does not contain \( e_1 \), since if \( e_1 \) were in \( F(e_1) \), then \( F(e_1) \) would be a proper subset of itself, which cannot be true. We may define \( P(e) \) as a function from elements to sets of elements: \( P(e) = (e \cup F(e)) \). In words, \( P(e) \) is “everything else except e and \( F(e) \)”. The union of \( P(e) \), e, and \( F(e) \) thus exhausts all the elements in the universe under discussion; further, they do not intersect. We may say
B: no element may both precede and follow e.

We may define classes of time other than Euclidean time by eliminating certain of these properties. Thus discrete time drops the continuity principle, but retains infinity, linearity, immediacy, universality, staticity, and unidirectionality. Finite time, of which linguistic time is an instance, further eliminates the requirement of infinity. One might also drop the requirement of staticity, so that events have duration of only one time unit.

Linear, linguistic representations have the character of non-static, finite, discrete time. Thus, for example, linear sequences of feature matrices, which that form phonological representations in SPE, have the properties of immediacy (the association of elements with particular locations in the ordering), universality (the claim that all elements are thus associated with some location), and linearity (the complete ordering of all the locations), but not of continuity, infinity, or staticity. Staticity is a special treatment in this case: geminate segments might be considered as states, since the same feature matrix repeats itself at two successive times. However, I take it as essential about SPE representations that repeating identical matrices are precisely that: repetitions; they are NOT the same matrix extended across two times. In this respect, SPE is unlike both its predecessors and its successors.

The linearity of time gives an order to temporal events/states/elements. In static finite-temporal representations, this ordering is represented by in some way adding to the description of each element of the sequence a uniquely ordered index. This is the implicit effect of writing forms in a sequence on paper: the physical layout provides an analog of the temporally ordered indexing of all the elements in the string. If this implicit ordering is done explicitly, by numbering the elements with an (ordered) index, then the actual location of elements in a particular physical representation becomes immaterial. Words in a sentence, for example, could be written on separate pieces of paper with their respective

that elements in P(e) are 'before' e, and elements in F(e) are 'after' e, though we could as well say that P(e) is after and F(e) is before e, and linearity would still be satisfied.

In discrete time staticity may be defined as the property that an event may occur at more than one time, s.t. the times at which it occurs form a "contiguous sequence". A set of times T forms a contiguous sequence iff there is no proper subset T1 of T for which T1 and T ∩ T1' (the complement of T1) have no "adjacent" members. Times t1 and t2 are adjacent iff there is no time t3 such that either t1 < t3 < t2 or t2 < t3 < t1, where "<" may be defined trivially using the functions F() and P(): t1 < t2 iff t1 ∈ P(t2).

As implied above, these terms are synonymous within this discussion.

290
indices and tossed into the air: their linear ordering is still fully specified since their (ordered) indices are specified. It makes no difference whether elements are concretely put in order (for example, by printing them in a spatial sequence on a teletype machine) or ordered abstractly by associating unique indices/times – for which the set of indices has the properties of finite time – with the elements (for example, by storing numbers with the elements in unordered, random-access, computer memory). In either case their ordering is fully specified; that is, concrete and abstract ordering are equivalent for our purposes.

This view of ordering adds no new descriptive machinery to linguistic representations. The features of finite time are implicit in any finite, linearly ordered, non-reversible sequence of discrete elements. Sequences of linguistic elements may therefore be described in terms of discrete time. If one wishes to add to a system of linguistic representation explicit machinery which indexes each element with a distinct “time” (or number, say), this makes no further claims than those already implicitly made by the unindexed, concrete representation of the same sequences. Thus this view merely makes explicit and formal the formerly implicit and informal machinery which has always been used in linguistic representations.

A.2 Space

Further indexing power is obtained if we add “space” to “time”. If only one event or element could occur at a time, as in considering any single time series, such as a waveform, the character output of a typewriter, the location of an automobile through the course of a day, etc., then mere linear ordering (for “normal time” this requires some fixed unit of measurement) of the events would be enough to orchestrate them completely. But if events may occur simultaneously, as when multiple time series are considered abstractly, another property of time becomes crucial: Each time serves as an index linking all the events (or states, or elements) occurring at that time. A many-to-one mapping between events and times is what is necessary for this more powerful indexing system, which we may call “multi-linearity”.

It is a further, perhaps unnecessary, stipulation that there are a fixed number of “places” each of which must index some element at each and every time. Some representations
may require an unstructured simultaneity of elements, rather than a fixed, ordered list of required elements. (This difference is analogous to the difference between variable-record-length signal representations and fixed-record-length signal representations).

In phonology, if the organization of sequences of phonological feature matrices is of the former, un-structured sort, then the features present at a given time (in addition to their values) may change. On the other hand, if such matrix-sequences are of the latter sort, then we may speak of fixed “places” or feature-locations, each of which, at every time, indexes some event, state or element — or in the phonological universe, some feature-value. In the case of a fixed set of time-series, we may speak of place-structured simultaneity; on the other hand if we have a different, arbitrary listing of elements for each time, we have unstructured simultaneity. One may imagine various degrees and forms of partly-structured simultaneity, as for example if there is a regular alternation between \( n \) sets of elements. Phonological constituents such as syllables, feet, phrases, etc., may be thought of as describing the structure of the alternation of feature inventories across time. Autosegmental tiers are simply locations in phonological space. In the rhythmic alternation across syllables or other phonological constituents, the inventory of features varies over constituent, so that, say, consonant features occur in onsets, vocalic ones in codas, etc. If features, or tiers, are simply locations in phonological space, then the alternation across syllables from one feature-inventory to another is a sort of rhythmic appearance and disappearance of the locations.

This shows limits to the space-and-time metaphor. As the vocal tract moves in speech production, different places become relevant to sound production at different times within syllable structure. Appearing and disappearing might rather be thought of as coming-into-relevance and falling-out-of-relevance. The different articulators that constitute the different “locations” in phonological space become relevant and irrelevant to sound production in association with syllable structure and other rhythmic constituents. This reflects the structure of physical actions in speech production as much as it reflects phonological structure. The patterns of rhythmic spatio-temporal structure in phonology is derived directly from a schematic view of the physical vocal tract. The tongue body features associated with vowels are not relevant to bilabial stop production, so those features can be considered nonexistent (or inactive) at the point of specifying the stop.
In any case, the task of orchestrating the events of multiple simultaneous time series requires a way of relating the events of one series to those of another: this implicitly requires what I here call finite-time-and-place indexing machinery, or multilinearity. Thus, to take a special case again, all events occurring now are linked in time; they “temporally overlap”, or “occur at the same time”, or “are indexed by the same time index”, and also all events occurring in just one of the time series “occur at the same place”, “spatially overlap”, or “are indexed by the same place index”. To summarize the universality property of multi-linear indexing in one predicate-calculus statement:

C: Each element which occurs is located at some place and at some time.

A.3 Linguistic Time

Linguistic time (LT) shares some characteristics of Euclidean time, namely immediacy, universality, linearity, staticity, and unidirectionality, but lacks the features of infinity and continuity.

First, LT is discrete rather than continuous – this derives from the discreteness of linguistic forms. How finely one wishes to divide the LT dimension for analytical or expository purposes is arbitrary up to a point: No finer linguistic temporal unit may be admitted than the smallest desired linguistic unit itself – the smallest unit may not be indexed by separate times. Indeed, beyond some level not much below that of the phoneme, linguistic time cannot be further subdivided. Normal time, on the other hand, is susceptible to much finer subdivision, to the point that it is not inappropriate to describe normal time as “continuous”.

Second, the LT dimension is finite, though unbounded, in the same way as are well-formed linguistic units like sentences or conversations. This follows from the preceding consideration: the number of times used to sequence the elements of a particular linguistic form cannot be greater than the number of smallest linguistic units in that form. Every particular linguistic form is finite (notwithstanding the claim that there is no upper bound to the length of possible sentences of given languages), and thus contains a finite number of smallest units, each of which may be indexed by at most one distinct time. For any
given form, then, the dimension of linguistic time is finite. Euclidean time on the other hand is infinite, and extends indefinitely into the past and future.

Linguistic time is also distinct from normal or Euclidean time in that the “same” linguistic event or object may be produced or comprehended in varying amounts of normal time. Linguistic representations may be seen extratemporally, in the sense that speakers (and linguists) may examine them in normal time, somehow outside their internal temporal sequence. The possible repeated mental reviewing of a sentence during the comprehension process is not different in this connection from rereading a sentence as a linguist, perhaps while considering it, usefully, as a fixed, extra-temporal, abstract structure. In both cases, a record (whether in memory or on paper) of a temporal sequence is reexamined after the events recorded. Any record of a temporal sequence, examined after the events recorded, is indeed a static abstract structure. This is the point linguists make by avoiding reference to time in linguistic theory. Still, they are “temporal” sequences nonetheless, not just because they are conceived, produced, perceived, and comprehended in normal time, but because linguistic representations have a specifically temporal organization.

Despite the differences between “real-time” and “linguistic time” there are enough similarities between them to justify retaining the term “time”, and associated temporal predicates such as “before” and “after”. The five remaining properties will be considered: immediacy, universality, staticity, linearity, and unidirectionality.

First, elements in linguistic representations occur at particular locations in the sequence. This property of elements of “occurring at” some (temporal) location is the same as having a special time, now, associated with the element. Linguistic representations have this feature, “immediacy”. Second, LT is “universal”. No element in a linguistic representation lacks a (temporal) location and extent. Every linguistic element in a representation is at some time. Third, linguistic representations in which elements are composed of a sequence of other elements (as phrases are composed of morphemes, morphemes of phonemes, etc.) have the feature of staticity. A constituent which contains elements at different temporal locations may be said to be “static” in the sense that that constituent extends across a span of times. To this extent LT is not exclusively linear, as in the phonological representations of SPE. Fourth, linear ordering is crucial in many linguistic representations: phonemes, words, phrases, and utterances are all formed into larger units.
by the device of linear ordering, by which a unit is constrained not to both precede and follow any other unit.

Fifth, and most important, linguistic representations are unidirectional. Despite this fact, most generative linguistics takes the opposite, bidirectionalist, and notably incorrect position. It is commonly assumed that "before" and "after" are arbitrarily chosen and might just as well be reversed. Thus there are left-branching as well as right-branching languages, languages with prepositional as well as postpositional phrases (for consistency with this bidirectionalist viewpoint, PP's should be termed "left-positional" and "right-positional" phrases), etc. The use of the metaphor of handedness in referring to temporal direction in linguistic representations would seem to emphasize the arbitrary nature of directionality in linguistic structure, since the direction of writing is widely understood to be arbitrary.

As noted before, the linearity of linguistic representations does not in itself require unidirectionality. Linearity is a formal property of strings of linguistic forms, while unidirectionality is an empirically established restriction on possible directions of ordering of elements in linguistic structures. Unidirectionality is an empirical finding, that a certain sequence can occur in one order, but not in the reverse order. In particular languages, of course, most syntactic and phonotactic structures are restricted to one direction of ordering or another, so within the discussion of a particular language, "before" and "after" are entirely appropriate. The issue is whether Universal Grammar has unidirectional phenomena.

In phonetics it is quite clear that ordering of events is unidirectional. A burst, for example, cannot precede its associated closure. Similarly, while jaw opening and closing may be mirror images of each other, each depends on the efforts of different, antagonistic muscles. So the muscular activities behind jaw movement cannot be described without reference to the direction of the movement in time.

Similarly vocal-fold vibration may depend on the preceding state of the glottis. There is an intermediate point of glottal opening at which the vocal folds will continue to vibrate if they had been vibrating immediately before, and at which the folds will not vibrate if they had not been vibrating immediately before, despite the fact that the larynx may have the same degree of glottal opening, vocal-fold tension, etc., in both cases. That is,
the voicing state is physically perseveratory, in this intermediate laryngeal position. This
effect is unidirectional; it cannot be anticipatory, since physical objects are not prescient.
Of course, such perseveratory physical effects must be distinguished from anticipatory
voicing assimilations, etc., which must be explained in other ways. Examples of temporal
irreversibility in phonetics may be multiplied indefinitely.

In phonology as well, there are unidirectional phenomena such as syllable-structure,
in which it is impossible for an onset to follow a rhyme, a coda to precede a nucleus,
etc. Similarly at the level of syntax, heavy-NP shift can only go in one direction, namely
forward.\(^5\) There may be many other linguistic phenomena which are unidirectional.

The formal marking of "linguistic time" may be done by a discrete, finite, unidirectional,
linearly ordered set of indices. The elements of any given representation are each
(universality) associated with (immediacy) one or more (staticity) indices. Linguistic time
thus constitutes a discrete, finite, linearly ordered dimension along which linguistic events
(or elements) are organized. "Times" within a linguistic representation may be thought of
either as formal objects associated with each element (in which case concrete ordering, such
as representing order between elements by their relations as written on paper, is obviated
) or as concrete "locations" of elements within a spatially or otherwise laid-out sequence
(which obviates formal marking of elements by indices).

The number of LT indices of a given sequence of linguistic forms is defined by the
number of elements in the sequence. Of course elements may be repeated in LT without
repeating their time of occurrence. In other words, there may be multiple tokens of a given
type.

Type-token confusions lead to logical traps which must be avoided. LT is an indexing
of the entire sequence in question. In examining a phrase in which a word repeats, the first
phoneme in the first instance of the word is not indexed by the same linguistic time as the
first phoneme in the second instance of that word. Of course, if the universe of discussion
is limited to that word, then the "first phoneme" has only one index. The formal indexing
of elements in a sequence depends on the sequence.

So far we have shown that certain features of time are implicit in linguistic represen-
tations. The game has been a formal one; we have not accounted for anything new that

\(^5\) Thanks to Richard Janda for this observation.
wasn't accounted for before; rather we have simply made explicit what was implicitly assumed in the past. The principles of linguistic time presented should be non-controversial, except for the issue of unidirectionality as applied to Universal Syntax. Now, we will see how the indexing power of linguistic "space", in conjunction with that of linguistic time, can be applied to autosegmental phonology.

Multi-linear phonology makes it necessary to add to the indexing power of LT additional features of structured simultaneity (or "place", in addition to time), in order to specify the organization of multiple "tiers" (sequences) of linguistic objects. This linking feature orchestrates multiple series of events or elements using both sequencing and overlapping.

A.4 Linguistic Space and the Well-Formedness Conditions

The Well-Formedness Condition of current multi-linear phonology includes three clauses, which we will discuss in the context of the relations between tones and vowels. In this context, the WFC states that:

1: For every vowel there is at least one tone.
2: For every tone there is at least one vowel.
3: Association lines may not cross.

These clauses are stated as though they are unrelated to each other and arbitrarily fixed; they call out for unification and explanation. They may be seen to follow trivially from features of linguistic spatio-temporal organization. Sagey (1986) gave a proof of the same conclusion using different assumptions, namely that phonological representations are continuous rather than discrete.

A.4.1 Proof of Well-Formedness Condition clauses

In addition to the given machinery of linguistic time, which is assumed for any linear representation of elements, if we additionally assume that there are two places, a vowel "place" (call it V), and a tone "place"(call it T), and we refer to all and only the elements occurring at V as vowels and we refer to all and only the elements occurring at T as tones,
then all three parts of the WFC follow. Multilinearity (condition C above) says that at each time there exists an element at each place. Therefore

D: At each time there is an element at V (a vowel).
E: At each time there is an element at T (a tone).

Proof of WFC #1: Choose an arbitrary vowel, call it v. By universality and immediacy, v occurs at some time, \( t_v \). But E states that there exists a tone at every time so there must a tone at \( t_v \). Therefore there is a tone at the same time as that vowel. Since the vowel was chosen arbitrarily, this is true for every vowel.

Proof of WFC #2: A symmetrical argument applies, with “tone” and “vowel” interchanged, and “D” replacing “E”.

Proof of WFC #3: Association lines in non-linear phonology correspond to simultaneous occurrence in an LT system. The corresponding claim made by clause 3 is that if an element \( v_1 \) precedes another element \( v_2 \), then elements cotemporal with \( v_1 \) and \( v_2 \), say \( t_1 \) and \( t_2 \), may not occur in the opposite order as their cotemporal elements. Consider the following propositions, i-iv, where \(<\) encodes “precedes” and \(\circ\) encodes association lines or simultaneity.

i: \( v_1 \circ t_1 \).
ii: \( v_2 \circ t_2 \).
iii: \( v_1 < v_2 \).
iv: \( t_2 < t_1 \).

Clause 3 of the Well-Formedness Condition essentially states that given i-iii, iv is ruled out. However, this follows trivially from fundamental assumptions about Linguistic time-and-place. By universality and immediacy, \( v_1 \) occurs at some time; let’s call it 1. By proposition i, \( t_1 \) is also at time 1. Similarly, \( v_2 \) and \( t_2 \) occur at some time, which we may call 2. Since \( v_1 \) precedes \( v_2 \), the time, 1, at which \( v_1 \) occurs, precedes the time, 2, at which \( v_2 \) occurs. But since \( t_1 \) follows \( t_2 \), the time at which \( t_1 \) occurs, 1, follows the time at which \( t_2 \) occurs, 2. But by linearity, one time may not both precede and follow another. Therefore 1 cannot both precede and follow 2, and one of our assumptions must be incorrect. To maintain propositions i-iii, we must abandon iv. QED.
The Well-Formedness Condition is at the very heart of non-linear phonology. The derivation of Clauses 1, 2, and 3 of the WFC from independent principles may therefore be considered an important achievement of an explicitly spatio-temporal perspective on linguistic representations.  

---

This paper further opens up questions of the cognitive representations of actual time which are not often frequently considered.
Appendix B

Historical Issues in Jamaican Creole

This appendix discusses two historical issues, relevant to the development of Jamaican Creole (JC). First, it discusses a functional relationship between long mid vowels and diphthongs, which may constitute a push chain shift. Second, the alleged historical phonological process of unmerger, which would seem to be necessary part of a certain outdated view of creole history, is defined and shown not to exist. Two classes of counterexamples (doublet borrowings in Arabic, and near-mergers) are argued to be false counterexamples. The empirical principle that mergers are irreversible is thereby validated. Finally, the consequences of this principle for theories of the historical development of creoles in general, and Jamaican Creole in particular, are outlined.

B.1 Sound shifts in Caribbean Creoles

Next we reconsider the historical sound changes which led to the relative positions of the vowels in Jamaican English Creole, which are quite different from the positions of corresponding vowels in the other English dialects studied in this thesis.

JC /e:, o:/ (as in FACE, GOAT) historically descend from Middle English (14th century) /a:, o/, which raised to Early Modern English /e:/ (circa 1650, merging with ME /e:/, according to Prins, 1972:122) and /o:/ (perhaps a century earlier). This raising
of ME long low vowels /a:, o:/ which occurred as a part of the Great Vowel Shift, was followed upon in other English dialects by diphthongization to [e'] and to [o* 8̃]. This diphthongization appears to postdate the establishment of Jamaican Creole: Prins states that “the Mo[dern]E[nglish] diphthongs [ei] and [ou] ... are late (end 18th c.).” (p. 124) In the antecedents of present-day Jamaican Creole, this process apparently did not occur. Rather, the subsequent step was a further raising of the nuclei of these vowels, so that they are now transcribed with /ie, uo/. While this raising is not strictly a part of the Great Vowel Shift, it can be seen as a continuation of the raising of these vowels which occurred in the GVS.1

Where did the glide in /ie, uo/ come from? Why did this change go to this extreme in JC. Notice that the raising of short-A (Wells' lexical set, TRAP) in American English dialects as far as [i9] is an example of the same raising process. It is clear that this identical vowel shift has occurred again and again in the history of English; presumably the same linguistic forces operated this century to raise “short A” in American dialects that operated in earlier JC to raise “long A” in a similar way.

One scenario to explain the further raising of /e:, o:/ relates this change to another special feature of Jamaican phonology. The sounds here labelled /ou, ai/ (as in MOUTH, PRICE) in Jamaican are the modern reflexes of Middle English /i:, u:/, which became the diphthongs /ay, aw/ of Reference American, via the Great Vowel Shift. In Jamaican, as in certain Northern dialects in Britain, the Great Vowel Shift has not gone to completion: the formerly high-back vowel in MOUTH has not shifted all the way down to low position. Certainly its nucleus is not to be identified with that of /ai/, as may be reasonably done in RA, as shown in Figure 6.5. Thus, MOUTH is /mo* t/.

The phonemes /ou/ and /o:/ (MOUTH and GOAT) would be quite difficult to distinguish without a phonetic difference in their respective nuclei: [o:] and [o*] are not phonetically very far apart. The quality of the diphthong /ou/ is mid-back and sometimes monophthongal, thereby putting functional pressure on the back phoneme /o:/, which is historically mid, back, and either monophthongal or up-gliding — that is, very much like /ou/. /o:/ might therefore have become differentiated from /ou/ by diphthongizing and

---

1 At the same time, other aspects of the GVS did not go as far in Jamaican Creole as elsewhere: the lowering of Middle English /u:/ (as in MOUTH, discussed above) did not go “to completion” as it did in other dialects.

301

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
raising to [uo]. The raising of /e:/ to [ie] could then appear as a parallel shift, a symmetric raising of a long mid vowel to a high, lowering diphthong.

It is useful to consider in this connection the range of realizations of the three relevant lexical sets in several varieties of Caribbean English. Wells presents phonological and phonetic summaries of several of these creoles, from which information in the following table was taken.

<table>
<thead>
<tr>
<th></th>
<th>FACE</th>
<th>GOAT</th>
<th>MOUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamaica</td>
<td>[ie]</td>
<td>[uo]</td>
<td>[ou]</td>
</tr>
<tr>
<td>Montserrat</td>
<td>[ie]</td>
<td>[uo]</td>
<td>[ou]</td>
</tr>
<tr>
<td>Trinidad</td>
<td>[e]</td>
<td>[o]</td>
<td>[ou]</td>
</tr>
<tr>
<td>Guyana</td>
<td>[e:]</td>
<td>[o:]</td>
<td>[ou]</td>
</tr>
<tr>
<td>Barbados</td>
<td>[e:]</td>
<td>[o:]</td>
<td>[u]</td>
</tr>
<tr>
<td>Bahamas</td>
<td>[e:]</td>
<td>[o:]</td>
<td>[u]</td>
</tr>
</tbody>
</table>

The sound written [ou] is not low or lower-mid, but mid or high-mid. In the above-described process, the falling of ME /u:/ (MOUTH) to [ou] is associated with the raising of /ow/(GOAT) to [uo], and the parallel raising of /e:/ (FACE) to /ie/. The only Caribbean creole dialect discussed by Wells with the nucleus of ME /u:/ in mid-back position (as [ou]) is that of Montserrat, one of the Windward Islands. The striking coincidence here is that Montserrat is also the only other dialect with raising and breaking of /e:, o:/ . This is evidence that the mid-back nucleus of the reflex of ME /u:/ and the raising and breaking of /e:, o:/ are systematically associated. The generalization here is functional:

Where ME /u:/ (MOUTH) becomes [ou], encroaching on /o:, /e:, o:/ raise and break.

This generalization supports the interpretation above of a push chain shift where /ou/ encroached on the phonetic space of /o:/, which then diphthongized and raised to high-back [uo], while /e:/ diphthongized and raised in a parallel and symmetric shift. If on the other hand the raising of /e:, o:/ to [ie, uo] occurred before ME /u:/ became [o^n] in Montserrat and Jamaica (that is, if the changes occurred as a drag chain shift rather than a push chain shift) then there would be no explanation for the raising of /e:, o:/ to [ie, uo].
The push chain scenario, on the other hand, gives a plausible explanation for the raising of /e:, o:/, namely the functional pressure of the phonetic approximation of ou/ with /o:/.

**Push chains**

Push chains are not as well accepted as drag chains; they suggest functional explanations for a class of sound change. In historical phonetic change, when the phonetic realization of one phoneme becomes very similar to that of another phoneme, three things could logically happen. First, they could merge. This is not necessarily — or perhaps at all — deterred by the “functional load” of the distinction (that is, the amount of homonymy that would result from merger). Mergers can occur even when a sound distinction carries a huge functional load, as shown by the merger in Middle-Indo-Iranian of all the strident fricatives, s, š, š, resulting in large numbers of homonymy, or of the merger in Early Modern English of EME /e:/ and /e:/ (as in meet and meat)(Prins 1972:122). The second possible consequence of phonetic approximation is near-merger, discussed in the next section, where sounds remain approximated to one another but do not merge. Finally, a push chain could occur, where one of the phonemes becomes distinguished by some new phonetic difference. Given the tolerance for both merger and near-merger, it is something of a mystery why push chains should occur. In fact, some theoreticians (King 1969) claim for independent theoretical reasons that they are impossible, though others (e.g., Labov, Yaeger, & Steiner 1972:212) accept their existence. But this mystery is no more unusual than the existence of both mergers and near-mergers: Why should approximating vowels merge or instead remain distinct, though approximated? The historical conditioning of merger versus unmerger is not known; it cannot therefore be surprising that the conditioning of the third alternative is not known either.

The explanation given here for the raising of the nuclei of /e:, o:/ to [ie, uo] involves a push chain, whereby /ou/ encroaches on /o:/ and then /o:/ becomes differentiated by a raised nucleus and a different glide direction. This explanation thus a small contribution to the discussion of push chains. If it is true, then all three results: merger, near-merger, and push-chain shifts can occur. Opponents of push-chains must provide a better explanation for this set of vowel shifts.

2Herold (1991) provides a detailed examination of the conditions of merger.
B.2 Decreolization and Merger

Complete decreolization, when mergers relative to the standard language have occurred in the historical basilect, assumes an impossible form of sound change: unmerger. This section discusses the irreversibility of merger in detail, and then discusses the consequences of this principle for the view of the development of Jamaican Creole and other "decreolized" languages. If two sound classes are merged in lower strata, but not in higher strata of a creole community, then those social levels that retain the distinction, must have existed from the earliest time in the development of the creole. The arguments presented here support recent arguments for social and linguistic diversity in the development of "decreolizing" creoles (Patrick 1991).

B.2.1 Unmerger is impossible

An unmerger of two formerly merged phonemes is a sound change that restores the former distinction. The distinction between two sounds, to recapitulate the discussion of page 42ff, involves not just two pronunciations, or sounds, but also two sets of words, where the first sound occurs in words of the first set, and the second occurs in the second set. An unmerger, in a dialect which merged two sounds, would therefore require the reintroduction of, on the one hand, the distinction between the two pronunciations, and on the other, the distinction between the two word classes. That is, if X and Y merge into one sound, then to unmerge them, the dialect must not merely acquire the two phonetic targets x and y: The two phonetic forms must also be associated with the original sets of words which originally contained them. That is, an unmerger requires that the descendant reflexes of merged phonemes A and B are sorted out again into the original word classes, so that the new pair of sounds is pronounced in the correct set of words.

Unmerger is impossible. That is, it does not occur. Counterexamples to the principle that mergers are irreversible are false. Besides near-mergers, which are discussed below, the only well-documented counterexample is discussed in Haeri (1991), from which the following discussion is taken.
False Unmerger

Abdel-Jawad (1981) claims that in Ammani Arabic, the Classical Arabic phoneme /q/ (a voiceless uvular stop, commonly called qaf) is now “unmerging” from the glottal stop, /ʔ/. The /q/ of Classical Arabic fully merged with the /ʔ/ before the end of the 14th century, though some believe this merger to have been complete a few centuries before that (Garbell, 1958, cited in Haeri, 1991). In modern Ammani, there is once again a distinction between qaf and glottal stop sounds. Further, the words that contain the qaf are those which had qaf in the Classical language. These facts are the basis for the claim that qaf and glottal stop have unmerged.

According to Haeri (1991), two main facts argue against this claim. First, the means by which /q/ has been re-introduced into the phonemic inventory of the urban dialects is through lexical borrowings. These borrowings in turn have occurred as a result of mass education in the Classical Arabic language. Thus many of the borrowed items form doublets with /ʔ/ words which descended directly from Ammani Arabic’s ancestor language — Classical Arabic, or something very closely related to it. Due to semantic shifts, there are a number of doublets in which clearly no un merger would be possible. Consider a couple of examples.

qarrar 'to decide'
ʔarrar 'to make someone confess'
qawi 'strong'
ʔawi 'very'

Thus qarrar descended directly from Classical Arabic into modern Ammani, changing its form to ʔarrar with the merger of qaf and glottal stop, and changing its meaning through historical semantic processes. The Classical Arabic form qarrar was then borrowed into Ammani, with the borrowed form retaining its original meaning, and containing the re-introduced /q/ sound.

For there to have been an unmerger, the historical qaf words which currently have glottal stops should have lost their glottal stops to qafs. They have not, since direct-descendant words like ʔarrar, ʔawi remain unaffected. Further, if direct-descendant words
were indeed identical with the borrowed lexical items, so that they "became" the new forms through un merger, then no semantic changes could have taken place in the course of several centuries. These are absurd conditions.

Both Abdel-Jawad and Haeri show that lexical items which include qaf are mostly restricted to formal, learned words. The increase in the incidence of such items is a result of the spread of mass education, and not the phonological un merger of formerly merged word classes.

There is a grain of truth to the claim that this is an un merger: /q/ does now exist in various modern Arabic languages, and it occurs only in words which had /q/ in Classical Arabic.

But this is not an un merger. The direct, lineal descendants of Classical /q/ and /ʔ/ words are still uniformly pronounced with /ʔ/; they have not been sorted out in the required way. Only the newly borrowed forms have /q/, a fact which is understandable on grounds of borrowing, without reference to an otherwise undocumented form of sound change.

Near-Merger

Another class of alleged un mergers are those which Labov ascribes to near-merger, in which two sounds get so phonetically close to each other that observers claim they are merged and even their speakers have trouble perceiving the distinction. Nonetheless the two classes are consistently distinguished acoustically (and therefore also in articulation), and children, whose phonetic perceptual abilities are perhaps better than adults, are able to acquire the distinction. Later phonetic shifts may modify these phonetic targets so that they no longer are so close to each other, and observers then may claim that the sounds have mysteriously un merged. However, they never had actually merged in the first place.

It may help dispel the common disbelief among linguists in the existence of near-mergers to describe an example. The best studied near-merger is that of ferry vs. furry in Philadelphian White English. For a class project some years ago, for example, I conducted a commutation test with a friend who was born and raised in Philadelphia. We tape-recorded him reading a randomized list of 13 tokens of the words ferry and furry. I backed the tape up to the 4th token, replayed the just-read list at a comfortable volume, and had my friend categorize his own pronunciations as "a boat" (ferry) or "an animal" (furry).
His performance was 5 correct and 5 incorrect, which is exactly the expected score on a forced-choice, two-answer test if the answers are generated randomly, by coin-toss. That is, he couldn’t tell which was which, in listening to his own speech. Later acoustic and impressionistic analysis of the tape showed that the vowels were in fact perfectly categorizable. A single line could be drawn on an F1-F2 chart to divide all tokens of one category from all tokens of the other. And I could teach myself to hear which was which, so that I could pass the test with a 100% correct score.

Thus the near-merger situation is this: speakers properly associate with two particular lexical sets a measurable phonetic difference which is audible to a trained phonetician. Despite this apparent phonemic difference, the same speakers are unable to correctly categorize the minimal pairs that they themselves produce.

The cognitive explanation for near-mergers is unknown at this time. However, I here propose an empirically testable explanation, based on the following two facts.

First, children are excellent impressionistic phoneticians; they acquire the most detailed features of phonetic production, which they must be able to hear, since they can imitate them (cf. Read, 1975, referred to in Kiparsky 1988). As maturation progresses, however, this perceptual ability apparently degrades. In fact, what it means to acquire a phonology is to a great extent simply learning not to be sensitive to non-distinctive differences; an overall loss of phonetic sensitivity is consistent with (though not logically implied by) this general effect.

Second, adults largely retain the phonetic patterns of their youth. In their youth, humans seem to learn an ingrained phonetic production system, precisely modeling the finest details of pronunciation and rendering them thoroughly automatic and unconscious. This pattern remains ingrained, crystallized, largely unmodified throughout the remainder of life. Studies of sound change in progress depend upon this fact in order to infer that the way teenagers were pronouncing things decades ago is much the same as the way the very same, now much older individuals pronounce things today. Thus a snapshot of the speech community at one moment in history that shows a pattern of age-grading in phonetic forms, could only reflect historical changes in progress if the people had this crystallized pattern of phonetic production.

These two facts may explain the near-merger situation. Children hear speech around
them in which some phonetic distinction occurs and is associated with particular sets of words; they learn to match the pronunciations and to use the right sound with the right words. But what happens when the children mature? First, the pattern of production has become ingrained and remains unchanged. And second, their perceptual abilities degrade. Some phonetic distinctions may quite well be so small that a degradation in perceptual capabilities could make it impossible for mature speakers to categorize the sound itself. Thus speakers could acquire a distinction, retain it in production, yet be unable to use it in perception.

Notice that children can learn such a distinction from adults that produce it, whether or not their adult models are able to perceive the distinction. Children learn language based on the productions of their models, not based on what their models can perceive! In this way a historically stable situation could occur in which two sounds are produced with a small phonetic difference, by an entire community, but where none of the adults are able to actually make linguistic, perceptual use of the distinction.

Experiments to test predictions of this hypothesis are quite feasible. For example, children should pass a categorization test while adults fail it. Such experiments are beyond the scope of this thesis; I hope to carry them out in future research.

The Possibility of Unmerger

The irreversibility of merger is an empirical principle, not a conceptual impossibility. In the right situation it could occur. A merger might be reversible under extreme conditions of dialect mixture, where speakers have sufficient exposure to (that is, nearly constant interaction with) speakers of a dialect where the merger did not take place, and further where the first speakers have strong motivation to learn the sound system of that dialect. This exposure must occur at a time during language acquisition when speakers are learning phonological categories and the lexical distributions of those categories. Further, the speakers of the unmerged dialect must not accommodate to the merged speakers by neutralizing the distinction in production, as was shown to occur under similar conditions of dialect contact by Herold (1990), in a study of the progress of the low-back merger in Pennsylvania. These conditions are sufficiently rare that they might well never be satisfied.

Mergers have always been found to expand at the expense of distinctions. This fact is
the foundation of the comparative method of historical linguistics: if two related languages are examined and a distinction is found in one that does not occur in the other, then the proto-language from which they descend is inferred to have had that distinction.

With considerable effort as well as constant exposure, isolated speakers, immersed among speakers of another dialect, may be able to acquire a phonetic distinction and its lexical distribution which was merged in his or her native dialect. No law of reason or nature says that merger is irreversible. However, it is a far-fetched possibility that an entire community of speakers should have this much exposure and motivation, at the right age, for this kind of sound change to occur. If an entire community were surrounded by speakers of another dialect, they wouldn't be surrounded, because there would be as many of the members as there are speakers of the other dialect. Further, in this kind of situation, the powerful forces of covert prestige are likely to exert themselves to prevent assimilation to the other dialect. Finally, imitation of other dialects can be "good enough" to pass as a successful imitation when only a relatively few phonetic targets are imitated; an imitation can well succeed without being a fully accurate one. So if the goal is successful imitation of a target dialect, it can be attained without going to the extreme of acquiring a new sub-lexicon.

In summary, there is no clear evidence in favor of, and considerable reason to doubt, the reversibility of mergers. Merger is irreversible.

B.2.2 Unmerger and Decreolization

What relevance do the above observations have to the theory of decreolization? Decreolization is hypothesized form of language shift in which parts of the community that formerly spoke a basilect, gradually acquire the forms of the acrolect. In complete decreolization, the basilect gradually disappears, and the population eventually comes to speak the standard language.

3For example, after several years as a graduate student in a relic area where a merger that occurs in my native dialect had not occurred (/ð u/), spending my time explicitly studying speech sounds, I may have acquired the distinction, at least in some words where spelling gives the correct clue. Thus I was recently unable to understand a low-back-merger speaker claiming someone was a [frıd], because I expected [b] in the word fraud and somehow couldn't understand what a frodd could possibly be. So insofar as I can be confused by the wrong phonetic form in a word, I may be functioning rather like a relic-area speaker.

4Cf. Graff, Labov, and Harris, 1986, and other papers in that group, which showed that blacks successfully imitating Philadelphia white speakers front /aw/ but not /aw/.

309
However, if mergers occurred in the process of creole formation, then direct descendants of the resulting merged dialect could never unmerge those sounds. If truly complete decreolization in fact occurred, then mergers would have to be reversed, so that the decreolized dialect is just like the (never-merged) standard. Otherwise there would remain differences between the speech descending from the creole and the standard to which it assimilates. Can we reconcile a view of creole development with the impossibility of unmergers? This discussion is inconclusive, with respect to Jamaican Creole, but it proposes a rule of inference from which, once the linguistic facts are clarified, the historical facts which resulted in them may be inferred.

Until recently, a commonly accepted view of the process of creole formation and development was as follows: At the time of the origin of the creole, a linguistic cataclysm occurred, and the deepest creole was formed. At this time the great majority of people in the community were speakers primarily of this basilect. Then in a lengthy historical process of assimilation, a range of intermediate lects emerged which are gradually more and more like the locally relevant standard language, until the continuum of lects emerged that is found today.

Consider the facts of Jamaican Creole, in this context. Notice that a number of lexical distributions are lost in the basilect. As discussed in Chapter 6, vowel of Reference American /ir, er/ (as in NEAR, SQUARE) are merged into the JC /ir/ class; /ur, or/ (as in CURE, FORCE) merged into JC /ur/; /ay, oy/ (as in PRICE and CHOICE) are merged into the /ai/ class; /æ, a/ (as in TRAP and LOT) are merged into the /a/ class, /ɔ:/ (THOUGHT), /ɑ:/ (BATII, PALM), /ar/ (START), /ɔr/ (NORTH, but not FORCE)/ are all merged into the /a:/ class. These mergers presumably occurred at the earliest stage in the development of the creole; they are characteristic of the basilect.

Mergers are irreversible. Therefore none of these mergers could be unmerged in any dialect descended from the original basilect. If acrolectal Jamaican Creole now has some of these distinctions, with the right words in the right sound-classes, then how is this compatible with the irreversibility of merger? Either decreolization goes, or irreversibility goes. We can’t have both, in this absolute form. If we maintain the principle of irreversibility of mergers, we must instead propose a different historical process. I therefore propose the following rule of inference for creole studies:

310
If historically accurate word classes are kept separate at one level of the creole continuum, but are merged at another level, then the level that which keeps them separate must have existed continuously throughout the history of the community.

If the acrolect has become so approximated to the standard, to the point that vowel classes are unmerged, then the reality must be that the unmerged acrolect has been there since the beginning.

Thus our understanding of the history of the acrolect depends on the facts of lexical distributions of sounds. Do any such apparent unmergers actually exist in situations of alleged decreolization? The essential question is the lexical distribution of sounds in the creole. J.C. Wells (1982:576) says that there are low-back vowels in acrolectal speech which can be pronounced in the formerly back-rounded word classes of LOT, THOUGHT, etc. But these classes are said to be fully merged in the basilect with word classes that don't allow the back-rounded vowel quality. I have difficulty believing that Wells is right, that the acrolectal speakers get their /ə/’s in the right words, and don’t put them in the wrong words. Unfortunately this matter will not be settled in this thesis. If Wells is correct, then the consequence is that the acrolect must have a long history, at least as long as the history of the community which now speaks it. This supports a much more finely graded view of the sociolinguistic situation in the early creole: the acrolect, without the basilect’s mergers, must have been present from the beginning.
Appendix C

Average Duration Measurements

Table C.1: Jamaican Creole (Juba): Vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>54</td>
<td>28</td>
<td>1</td>
<td>496</td>
</tr>
<tr>
<td>/u/</td>
<td>58</td>
<td>31</td>
<td>3</td>
<td>139</td>
</tr>
<tr>
<td>/e/</td>
<td>83</td>
<td>42</td>
<td>3</td>
<td>249</td>
</tr>
<tr>
<td>/o/</td>
<td>91</td>
<td>48</td>
<td>3</td>
<td>297</td>
</tr>
<tr>
<td>/a/</td>
<td>87</td>
<td>41</td>
<td>2</td>
<td>694</td>
</tr>
<tr>
<td>/e/</td>
<td>147</td>
<td>56</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>/ii/</td>
<td>116</td>
<td>48</td>
<td>5</td>
<td>91</td>
</tr>
<tr>
<td>/uu/</td>
<td>133</td>
<td>45</td>
<td>6</td>
<td>52</td>
</tr>
<tr>
<td>/ie/</td>
<td>135</td>
<td>63</td>
<td>8</td>
<td>69</td>
</tr>
<tr>
<td>/uo/</td>
<td>122</td>
<td>47</td>
<td>4</td>
<td>144</td>
</tr>
<tr>
<td>/ou/</td>
<td>156</td>
<td>73</td>
<td>7</td>
<td>102</td>
</tr>
<tr>
<td>/aa/</td>
<td>190</td>
<td>70</td>
<td>6</td>
<td>119</td>
</tr>
<tr>
<td>/ai/</td>
<td>146</td>
<td>52</td>
<td>4</td>
<td>191</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>76</td>
<td>41</td>
<td>1</td>
<td>1875</td>
</tr>
<tr>
<td>Long Vowels</td>
<td>144</td>
<td>62</td>
<td>2</td>
<td>768</td>
</tr>
<tr>
<td>Stressed</td>
<td>136</td>
<td>64</td>
<td>2</td>
<td>914</td>
</tr>
<tr>
<td>Unstressed</td>
<td>76</td>
<td>41</td>
<td>1</td>
<td>1766</td>
</tr>
<tr>
<td>/_C[+voice]</td>
<td>103</td>
<td>56</td>
<td>2</td>
<td>1087</td>
</tr>
<tr>
<td>/_C[-voice]</td>
<td>95</td>
<td>56</td>
<td>2</td>
<td>632</td>
</tr>
<tr>
<td>/C[+voice]_</td>
<td>99</td>
<td>57</td>
<td>2</td>
<td>1198</td>
</tr>
<tr>
<td>/C[-voice]_</td>
<td>98</td>
<td>62</td>
<td>3</td>
<td>554</td>
</tr>
</tbody>
</table>
Table C.2: Jamaican Creole (Roasta): Vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>54</td>
<td>24</td>
<td>1</td>
<td>433</td>
</tr>
<tr>
<td>/u/</td>
<td>57</td>
<td>19</td>
<td>2</td>
<td>107</td>
</tr>
<tr>
<td>/e/</td>
<td>63</td>
<td>21</td>
<td>1</td>
<td>199</td>
</tr>
<tr>
<td>/o/</td>
<td>63</td>
<td>22</td>
<td>2</td>
<td>108</td>
</tr>
<tr>
<td>/a/</td>
<td>81</td>
<td>25</td>
<td>1</td>
<td>341</td>
</tr>
<tr>
<td>/l/</td>
<td>60</td>
<td>36</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>/σ/</td>
<td>147</td>
<td>40</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>/ii/</td>
<td>96</td>
<td>34</td>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>/uu/</td>
<td>108</td>
<td>36</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>/ir/</td>
<td>118</td>
<td>28</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>/ur/</td>
<td>105</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>/ie/</td>
<td>96</td>
<td>29</td>
<td>4</td>
<td>46</td>
</tr>
<tr>
<td>/uo/</td>
<td>102</td>
<td>47</td>
<td>7</td>
<td>52</td>
</tr>
<tr>
<td>/ou/</td>
<td>139</td>
<td>55</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>/aa/</td>
<td>150</td>
<td>39</td>
<td>4</td>
<td>101</td>
</tr>
<tr>
<td>/ar/</td>
<td>142</td>
<td>45</td>
<td>5</td>
<td>69</td>
</tr>
<tr>
<td>/ai/</td>
<td>134</td>
<td>43</td>
<td>4</td>
<td>97</td>
</tr>
</tbody>
</table>

Short Vowels 64 25 1 1188
Long Vowels 125 46 2 390
Stressed 120 50 2 460
Unstressed 68 30 1 1213
/—C[+voice] 90 41 2 559
/—C[-voice] 87 48 3 320
/[+voice] 83 41 2 449
/[−voice] 78 44 3 260
Table C.3: Chicago White English (Rita): Vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>66</td>
<td>44</td>
<td>5</td>
<td>88</td>
</tr>
<tr>
<td>/I/</td>
<td>69</td>
<td>34</td>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>/U/</td>
<td>93</td>
<td>38</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>/e/</td>
<td>105</td>
<td>55</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td>/æ/</td>
<td>102</td>
<td>52</td>
<td>6</td>
<td>75</td>
</tr>
<tr>
<td>/œ/</td>
<td>107</td>
<td>54</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td>/a/</td>
<td>144</td>
<td>60</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>/I/</td>
<td>41</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>/œ/</td>
<td>112</td>
<td>39</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>/i/</td>
<td>79</td>
<td>35</td>
<td>4</td>
<td>77</td>
</tr>
<tr>
<td>/u/</td>
<td>73</td>
<td>41</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>/e/</td>
<td>103</td>
<td>30</td>
<td>5</td>
<td>39</td>
</tr>
<tr>
<td>/ow/</td>
<td>136</td>
<td>57</td>
<td>10</td>
<td>31</td>
</tr>
<tr>
<td>/oy/</td>
<td>180</td>
<td>78</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>/ay/</td>
<td>116</td>
<td>37</td>
<td>4</td>
<td>77</td>
</tr>
<tr>
<td>/aw/</td>
<td>135</td>
<td>41</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>/œ/</td>
<td>176</td>
<td>82</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>91</td>
<td>53</td>
<td>3</td>
<td>376</td>
</tr>
<tr>
<td>Long Vowels</td>
<td>107</td>
<td>50</td>
<td>3</td>
<td>309</td>
</tr>
<tr>
<td>Stressed</td>
<td>127</td>
<td>57</td>
<td>3</td>
<td>315</td>
</tr>
<tr>
<td>Unstressed</td>
<td>74</td>
<td>32</td>
<td>2</td>
<td>371</td>
</tr>
<tr>
<td>/C[+voice]</td>
<td>100</td>
<td>55</td>
<td>3</td>
<td>341</td>
</tr>
<tr>
<td>/C[-voice]</td>
<td>99</td>
<td>45</td>
<td>4</td>
<td>147</td>
</tr>
<tr>
<td>/C[+voice]</td>
<td>103</td>
<td>48</td>
<td>3</td>
<td>283</td>
</tr>
<tr>
<td>/C[-voice]</td>
<td>86</td>
<td>47</td>
<td>4</td>
<td>154</td>
</tr>
</tbody>
</table>
Table C.4: Chicago White English (Judy): Vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>60</td>
<td>25</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>/i/</td>
<td>131</td>
<td>86</td>
<td>9</td>
<td>92</td>
</tr>
<tr>
<td>/u/</td>
<td>134</td>
<td>74</td>
<td>7</td>
<td>102</td>
</tr>
<tr>
<td>/ə/</td>
<td>118</td>
<td>75</td>
<td>8</td>
<td>95</td>
</tr>
<tr>
<td>/ɑ/</td>
<td>130</td>
<td>58</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>113</td>
<td>58</td>
<td>5</td>
<td>113</td>
</tr>
<tr>
<td>/ɨ/</td>
<td>96</td>
<td>55</td>
<td>4</td>
<td>154</td>
</tr>
<tr>
<td>/ɜ/</td>
<td>255</td>
<td>125</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>119</td>
<td>104</td>
<td>18</td>
<td>33</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>132</td>
<td>83</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>145</td>
<td>63</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>/ɔɪ/</td>
<td>133</td>
<td>66</td>
<td>6</td>
<td>117</td>
</tr>
<tr>
<td>/oɪ/</td>
<td>190</td>
<td>56</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>/oɪ/</td>
<td>173</td>
<td>69</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>159</td>
<td>47</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td>/ɑɾ/</td>
<td>173</td>
<td>90</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>122</td>
<td>70</td>
<td>5</td>
<td>177</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>170</td>
<td>81</td>
<td>13</td>
<td>39</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>90</td>
<td>66</td>
<td>2</td>
<td>823</td>
</tr>
<tr>
<td>Long Vowels</td>
<td>126</td>
<td>73</td>
<td>3</td>
<td>792</td>
</tr>
<tr>
<td>Stressed</td>
<td>148</td>
<td>79</td>
<td>3</td>
<td>745</td>
</tr>
<tr>
<td>Unstressed</td>
<td>73</td>
<td>40</td>
<td>1</td>
<td>870</td>
</tr>
<tr>
<td>/_C[+voice]</td>
<td>111</td>
<td>73</td>
<td>3</td>
<td>641</td>
</tr>
<tr>
<td>/_C[-voice]</td>
<td>101</td>
<td>60</td>
<td>3</td>
<td>387</td>
</tr>
<tr>
<td>/C[+voice]_</td>
<td>114</td>
<td>74</td>
<td>3</td>
<td>668</td>
</tr>
<tr>
<td>/C[-voice]_</td>
<td>103</td>
<td>65</td>
<td>3</td>
<td>401</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table C.5: Chicago White English (Jim): Vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>57</td>
<td>20</td>
<td>1</td>
<td>505</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>71</td>
<td>28</td>
<td>2</td>
<td>259</td>
</tr>
<tr>
<td>/u/</td>
<td>76</td>
<td>26</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>103</td>
<td>46</td>
<td>4</td>
<td>108</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>115</td>
<td>68</td>
<td>6</td>
<td>146</td>
</tr>
<tr>
<td>/æ/</td>
<td>114</td>
<td>49</td>
<td>4</td>
<td>150</td>
</tr>
<tr>
<td>/æ/</td>
<td>110</td>
<td>40</td>
<td>4</td>
<td>83</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>87</td>
<td>34</td>
<td>3</td>
<td>94</td>
</tr>
<tr>
<td>/ɪr/</td>
<td>131</td>
<td>47</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>/iy/</td>
<td>76</td>
<td>28</td>
<td>2</td>
<td>155</td>
</tr>
<tr>
<td>/uw/</td>
<td>96</td>
<td>34</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>/ey/</td>
<td>98</td>
<td>33</td>
<td>4</td>
<td>82</td>
</tr>
<tr>
<td>/ər/</td>
<td>113</td>
<td>57</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>/ow/</td>
<td>127</td>
<td>53</td>
<td>4</td>
<td>147</td>
</tr>
<tr>
<td>/or/</td>
<td>151</td>
<td>68</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>/oy/</td>
<td>245</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>/o/</td>
<td>149</td>
<td>49</td>
<td>7</td>
<td>51</td>
</tr>
<tr>
<td>/ar/</td>
<td>134</td>
<td>41</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>/ay/</td>
<td>110</td>
<td>41</td>
<td>2</td>
<td>325</td>
</tr>
<tr>
<td>/aw/</td>
<td>133</td>
<td>45</td>
<td>6</td>
<td>62</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>81</td>
<td>45</td>
<td>1</td>
<td>1296</td>
</tr>
<tr>
<td>Long Vowels</td>
<td>109</td>
<td>47</td>
<td>1</td>
<td>1033</td>
</tr>
<tr>
<td>Stressed</td>
<td>123</td>
<td>52</td>
<td>2</td>
<td>941</td>
</tr>
<tr>
<td>Unstressed</td>
<td>73</td>
<td>32</td>
<td>1</td>
<td>1388</td>
</tr>
<tr>
<td>/C[+voice]</td>
<td>92</td>
<td>51</td>
<td>2</td>
<td>877</td>
</tr>
<tr>
<td>/C[-voice]</td>
<td>94</td>
<td>38</td>
<td>2</td>
<td>557</td>
</tr>
<tr>
<td>/C[+voice]_</td>
<td>101</td>
<td>49</td>
<td>2</td>
<td>859</td>
</tr>
<tr>
<td>/C[-voice]_</td>
<td>89</td>
<td>44</td>
<td>2</td>
<td>501</td>
</tr>
</tbody>
</table>
Table C.6: Anniston, Alabama English (James H.): Vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ə/</td>
<td>69</td>
<td>30</td>
<td>2</td>
<td>171</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>94</td>
<td>45</td>
<td>3</td>
<td>238</td>
</tr>
<tr>
<td>/u/</td>
<td>93</td>
<td>43</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>139</td>
<td>65</td>
<td>5</td>
<td>156</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>104</td>
<td>40</td>
<td>3</td>
<td>141</td>
</tr>
<tr>
<td>/æ/</td>
<td>133</td>
<td>42</td>
<td>4</td>
<td>108</td>
</tr>
<tr>
<td>/ɑ/</td>
<td>129</td>
<td>62</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>/l/</td>
<td>85</td>
<td>32</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>105</td>
<td>60</td>
<td>7</td>
<td>77</td>
</tr>
<tr>
<td>/i/</td>
<td>134</td>
<td>51</td>
<td>7</td>
<td>62</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>156</td>
<td>51</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>/uw/</td>
<td>120</td>
<td>52</td>
<td>6</td>
<td>75</td>
</tr>
<tr>
<td>/e/</td>
<td>130</td>
<td>50</td>
<td>5</td>
<td>105</td>
</tr>
<tr>
<td>/er/</td>
<td>168</td>
<td>62</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>/ow/</td>
<td>168</td>
<td>63</td>
<td>6</td>
<td>111</td>
</tr>
<tr>
<td>/or/</td>
<td>164</td>
<td>51</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>/oy/</td>
<td>333</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>165</td>
<td>52</td>
<td>6</td>
<td>73</td>
</tr>
<tr>
<td>/ɔr/</td>
<td>171</td>
<td>30</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>/ar/</td>
<td>192</td>
<td>48</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>/ay/</td>
<td>149</td>
<td>54</td>
<td>5</td>
<td>103</td>
</tr>
<tr>
<td>/aw/</td>
<td>189</td>
<td>45</td>
<td>6</td>
<td>65</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>105</td>
<td>53</td>
<td>2</td>
<td>858</td>
</tr>
<tr>
<td>Long Vowels</td>
<td>148</td>
<td>60</td>
<td>2</td>
<td>763</td>
</tr>
<tr>
<td>Stressed</td>
<td>154</td>
<td>58</td>
<td>2</td>
<td>886</td>
</tr>
<tr>
<td>Unstressed</td>
<td>91</td>
<td>42</td>
<td>2</td>
<td>751</td>
</tr>
<tr>
<td>/-C[+voice]</td>
<td>131</td>
<td>61</td>
<td>2</td>
<td>698</td>
</tr>
<tr>
<td>/-C[-voice]</td>
<td>113</td>
<td>45</td>
<td>2</td>
<td>375</td>
</tr>
<tr>
<td>/C[+voice]_</td>
<td>132</td>
<td>62</td>
<td>2</td>
<td>652</td>
</tr>
<tr>
<td>/C[-voice]_</td>
<td>114</td>
<td>55</td>
<td>3</td>
<td>430</td>
</tr>
</tbody>
</table>

317
Table C.7: Los Angeles Chicano English (Vince): vowel durations (ms)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/α/</td>
<td>59</td>
<td>33</td>
<td>2</td>
<td>318</td>
</tr>
<tr>
<td>/i/</td>
<td>71</td>
<td>37</td>
<td>3</td>
<td>192</td>
</tr>
<tr>
<td>/u/</td>
<td>67</td>
<td>34</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>92</td>
<td>42</td>
<td>4</td>
<td>109</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>114</td>
<td>67</td>
<td>6</td>
<td>123</td>
</tr>
<tr>
<td>/æ/</td>
<td>91</td>
<td>35</td>
<td>3</td>
<td>173</td>
</tr>
<tr>
<td>/l/</td>
<td>58</td>
<td>21</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>/œ/</td>
<td>79</td>
<td>40</td>
<td>4</td>
<td>106</td>
</tr>
<tr>
<td>/y/</td>
<td>74</td>
<td>40</td>
<td>3</td>
<td>210</td>
</tr>
<tr>
<td>/ir/</td>
<td>119</td>
<td>47</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>/uw/</td>
<td>100</td>
<td>56</td>
<td>8</td>
<td>46</td>
</tr>
<tr>
<td>/ey/</td>
<td>113</td>
<td>50</td>
<td>6</td>
<td>79</td>
</tr>
<tr>
<td>/er/</td>
<td>101</td>
<td>37</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>/ow/</td>
<td>129</td>
<td>56</td>
<td>5</td>
<td>111</td>
</tr>
<tr>
<td>/or/</td>
<td>140</td>
<td>80</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>/oy/</td>
<td>210</td>
<td>89</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>/a/</td>
<td>118</td>
<td>46</td>
<td>6</td>
<td>70</td>
</tr>
<tr>
<td>/ar/</td>
<td>129</td>
<td>37</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>/ay/</td>
<td>104</td>
<td>50</td>
<td>4</td>
<td>183</td>
</tr>
<tr>
<td>/aw/</td>
<td>153</td>
<td>78</td>
<td>17</td>
<td>22</td>
</tr>
</tbody>
</table>

**Short Vowels** 79 45 1 937
**Long Vowels** 102 54 2 903
**Stressed** 115 56 2 608
**Unstressed** 75 40 1 1192
/ /C[+ voice] 91 51 2 746
/ /C[-voice] 87 37 2 400
/ /C[+ voice] 92 50 2 829
/ /C[-voice] 88 55 3 442

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Bibliography


Eckert, Penelope. 1986. The roles of high school social structure in phonological change. Chicago Linguistics Society.


in press. "Short-A Raising in Southern States English?"


Fujimura, Osamu; Sproat, Richard. 1990. Evidence for the Non-Categorical nature of the clear-l, dark-l distinction. LSA paper???


323


Cambridge, Mass: MIT Press.


Cambridge, New York: Cambridge University Press.


Hindle, Donald M. 1979. The social and situational conditioning of phonetic variation.


Jakobson, Roman; Fant, Gunnar; Halle, Morris. 1952. Preliminaries to Speech Analysis.
Cambridge, Mass: MIT Press.


324


325

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


326


329

Rogers, Catherine. 1940. A dialect study of Camp Hill, Tallapoosa County Alabama. Alabama Polytechnic Institute Master's thesis.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


