PHONETIC DISTINCTIVENESS AS A SOCIOLINGUISTIC VARIABLE

by

Kevin Michael Heffernan

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Linguistics

University of Toronto

© by Kevin Michael Heffernan (2007)
I begin with a distinction between engendered variation, that is, sex-based phonetic-level variation that shows consistent patterns across communities, and nongendered variation. Much of engendered variation is anatomically-determined, but there is also an important behaviourally-determined component. One example of behaviourally-determined engendered variation is that women speak more clearly, as reflected in their more-dispersed vowel space. I argue that this sex-based difference applies to all types of phonological contrasts, not just vowels, so that the temporal and acoustic correlates of phonological constituents produced by women are differentiated typically more than those produced by men. Furthermore, I demonstrate that “phonetic distinctiveness” indexes other social categories such as social gender and social class, i.e., is a sociolinguistic variable.

I corroborate the indexing of social categories in two major applications. First, I compare phonetic correlates in the speech of eight male radio DJs from a range of music genres with ratings of how macho-sounding each DJ is. For each DJ, I examine durational differences between contextual and inherent long and short vowels. As expected, social gender significantly correlates with the vowel duration distinctiveness so that the more macho-sounding DJs produce less distinct vowel length contrasts.

Second, I examine dispersion of front and back vowels in the speech of 439
speakers from the *Atlas of North American English* (Labov et al. 2006). Vowel dispersion is shown to correlate with both sex and social class.

That men typically produce less distinct contrasts than women leads to predictions about the role of sex in phonetic-level sound change. Specifically, if the innovative form results in the loss of phonetic distinctiveness (such as two vowels merging), then we predict that men lead the change. An investigation of vowel mergers among the *Atlas of North American English* speakers reveals that men do lead mergers, and that speakers with a less dispersed vowel system show more instances of mergers, regardless of sex. Thus phonetic distinctiveness as an explanation of sound change accounts for why men lead sound changes that result in loss of phonetic distinctiveness, while women lead sound changes that maintain phonetic distinctiveness.
Acknowledgements

As I was nearing the completion of my dissertation, I read the article by Jacewicz et al. titled “Prosodic prominence effects on vowels in chain shifts.” In that article the authors conclude that women lead chain shifts by producing longer and clearer vowel variants, i.e., their more dispersed vowels. That study nicely complements my dissertation, in which I argue that men lead vowel mergers because of their reduced vowel space, i.e., their less clear vowels. Their independent discovery of very similar results was a source of inspiration and reassurance that my efforts were not in vain.

I would like to thank each of my committee members for the essential role they played from inception to completion of the dissertation. My initial interest in phonetic-level sex-based differences was sparked by Ron Smyth’s research on the phonetic cues to gay-sounding voices. Ron was also of invaluable assistance in designing the perception experiment, and running the statistics. I have doubled my knowledge of statistics, and yet, I still feel that I have just begun learning.

In my first year of my program, I took a self-directed course with Laura Colantoni. During that course, I read much of the phonetics literature that was reviewed in the second and third chapters. One of my more frustrating weeks of the dissertation process was trying to devise a method of measuring phonetic voicing based on the first derivative of the speech signal. It was Laura who saved me from further grief with the elegant, yet simple solution of zero-crossing rate.

It was also during that course that I discovered Jane Stuart-Smith’s work on /s/ in Glasgow, and her claim that sex-based anatomical differences provide a frame for the indexation of gender differences in the production of /s/. This claim led my to the
formation of the engendered / non-gendered distinction presented in Chapter One. As such, I was delighted that Jane was able to serve as the external. Jane’s comments and discussion were a perfect balance of praise and criticism. As I told Jack, as I was reflecting on them I could feel my ideas making the transition from initial proposition to critical reflection.

Becky Roeder joined the committee after the first four chapters had settled and begun to gather dust. As such, her appearance was a refreshing opportunity to rework a few of the more confusing points in these chapters. Interestingly, Becky was the most critical of the hypotheses, and I ended up abandoning one after she made me realize it was unattainable.

I am most indebted to my supervisor, Jack Chambers. Jack motivated me to start writing the thesis by pointing to my hand-drawn sketch of Figure 1 and saying that it needed to go on page one. Jack also motivated me to finish writing in a timely manner by giving me a hard deadline. I would also like to thank Jack for taking on the role of editor, and for show patience with my many, often repeated mutilations of English spelling and grammar. I am a better writer thanks to Jack.

All of my committee members helped make my experience as a graduate student a very positive one, and I am looking forward to continuing to develop our professional relationships in the years to come.
# Contents

1 Introduction: Two Types of Phonetic-level Variation ......................... 1  
   1.1 Introduction .............................................................................................. 1  
   1.2 The phonetic-level engendered variation framework .......................... 2  
      1.2.1 Two types of phonetic-level variation ........................................ 2  
      1.2.2 Extending the source of phonetic-level engendered variation ....... 6  
   1.3 Prior sociolinguistic studies of phonetic variation .............................. 6  
   1.4 Prior phonetic studies of social variation ............................................. 9  
   1.5 The significance of phonetic-level engendered variation ................... 12  

2 Anatomical Sources of Engendered Variation ............................ 15  
   2.1 What is engendered variation? ................................................................. 15  
   2.2 Voice Onset Time and Voicing ............................................................... 16  
      2.2.1 Variation of VOT by language and dialect ................................. 16  
      2.2.2 Variation of VOT and voicing by sex ......................................... 18  
      2.2.3 Perception of VOT ....................................................................... 23  
   2.3 Sibilant centre of gravity ........................................................................ 25  
      2.3.1 Centre of gravity studies of /s/ ...................................................... 25  
      2.3.2 Centre of gravity studies of /ʃ/ ..................................................... 31  
   2.4 Pitch ........................................................................................................ 34  
      2.4.1 Pitch as a marker of gender ............................................................ 34  
      2.4.2 Pitch as a marker of other social categories ................................. 39  
      2.4.3 Pitch as a marker of politeness ...................................................... 43  
   2.5 Breathiness .............................................................................................. 45  
      2.5.1 Breathiness as phonetic-level engendered variation .................... 45  
      2.5.2 Other studies of breathiness ......................................................... 48  
   2.6 Creak ........................................................................................................ 50  
      2.6.1 Creak as engendered phonetic-level variation .............................. 50
# Table of Contents

7.1 Revisiting the engendered–nongendered distinction .............................................. 215  
7.1.1 The engendered–nongendered distinction .................................................... 215  
7.1.2 Phonetic distinctiveness as a sociolinguistic variable ................................. 216  
7.1.3 Phonetic distinctiveness and language change ........................................... 218  

7.2 From specifics to generalizations, and back to specifics .................................. 221  
7.2.1 From specifics to generalizations: The first step ....................................... 221  
7.2.2 From generalization back to specifics: The next step ............................... 222  

Appendix A: List of Hypotheses ............................................................................ 225  
Appendix B: Transcriptions of Speech Samples .................................................. 228  
Appendix C: Zero-crossing Rate PRAAT Script ................................................. 233  
Appendix D: Merger Cross Tabulations .............................................................. 235  
References ............................................................................................................ 242
List of Figures

Figure 1.1 Two types of phonetic-level variation .......................................................3
Figure 2.1 Mean VOTs for velar plosives from 18 languages .........................17
Figure 2.2 Mean VOT voiced / voiceless difference for the bilabial and alveolar plosives by age and sex ..............................................................21
Figure 2.3 Variants of (s) by ethnicity in New Zealand ............................................29
Figure 2.4 /s/ peak frequency for eight groups of Glaswegians .........................30
Figure 2.5 Mean values and standard deviation bars for /s/ in two languages by sex .......................................................................................................31
Figure 2.6 Sibilant mean centre of gravity in Hz at 100 ms before voice onset before high vowels by age ..............................................................32
Figure 2.7 Schematic of a glottal cycle showing various measurements ...............52
Figure 2.8 A typical volume-velocity waveform for a male speaker ..................53
Figure 2.9 Single- and double-peak patterns during the production of creak (glottal fry) ...............................................................................................54
Figure 3.1 Scaling factors for male / female vowel normalization for seven languages and dialects ..............................................................................63
Figure 3.2 Mean formant values of vowels for British English Received Pronunciation, plotted in F1–F2 space ..............................................................65
Figure 3.3 Ellipses showing two standard deviations for the variation of the first and second formants of West Coast American English /a/ and /u/ for five men and five women .....................................................74
Figure 3.4 Average vowel duration for phonemically short and long vowels in Creek by sex ...............................................................................................77
Figure 3.5 VOT voiced / voiceless differences for high vowels and non-high vowels in Tokyo Japanese by age and sex .............................................78
Figure 3.6a Fronting of Canadian Raising nuclei by age and sex ......................81
Figure 3.6b  Rounding of Canadian Raising nuclei by age and sex.........................81

Figure 4.1  Two hypothetical examples illustrating the second Social Class Hypothesis .................................................................94

Figure 4.2  An example of variation of (r) by speech style and social class..........98

Figure 4.3  Progress of the COT–CAUGHT merger in Pottsville, Pennsylvania .................................................................111

Figure 4.4  Use of glottalized variant of (t) in Tyneside English in word-medial position in two speech styles by sex and social class.......113

Figure 4.5  Percentage of word-final pre-pausal tokens in word list style of /t/ that are the continued voicing variant and the extended frication variant in Newcastle .........................................................114

Figure 4.6  Vowel quality for the word “the” by following environment.........117

Figure 4.7  Use of [i] in “the” before words beginning with vowels by speaker age..................................................................................118

Figure 4.8  Percentage /ʒ/ devoiced in Buenos Aires Spanish by sex.............120

Figure 4.9  Index scores for 74 speakers of Alexandrian Arabic by speaker group .......................................................................................126

Figure 5.1  Percentage usage of the non-standard variant of the present participle verb suffix -ing by the eight DJs..............................................138

Figure 5.2  A sound wave showing the transition from a vowel to a fricative......152

Figure 5.3a  Zero crossings for a voiceless alveolar fricative.................................154

Figure 5.3b  Zero crossings for a voiced alveolar fricative....................................154

Figure 5.4  Average vowel durations before a voiced and voiceless consonants for the eight DJs in two accent environments, heavy and light.........................................................................................171

Figure 5.5  Average vowel durations for tense and lax vowels for the eight DJs in two accent environments, heavy and light......................172

Figure 5.6  The interconnectivity of our social gender and the phonetic realization of phonological contrasts ..................................................174

Figure 6.1  Scatterplots of Front Vowel Distinctiveness and Socio-Economic Index for the males and females ..................................................190

Figure 6.2  Scatterplots of Back Vowel Distinctiveness and Socio-Economic Index for the males and females .............................................194

Figure 6.3  The percentage of speakers with merged COT–CAUGHT in perception and production by region ....................................................204
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Percentage of fully voiced /h/ by sex</td>
<td>11</td>
</tr>
<tr>
<td>2.1a</td>
<td>Proportions of voiced and devoiced obstruents by region and sex in word-initial position</td>
<td>24</td>
</tr>
<tr>
<td>2.1b</td>
<td>Proportions of voiced and devoiced obstruents by region and sex in word-final position</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>Age, height, weight and speaking fundamental frequency data for two groups of American males</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary of some of the statistical results reported by Byrd (1994) in his examination of reduction in the TIMIT database</td>
<td>72</td>
</tr>
<tr>
<td>3.2</td>
<td>Average first and second formant values and emphasis for /a/ by preceding consonant</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>Number of errors made in the identification of children’s sex by social class</td>
<td>90</td>
</tr>
<tr>
<td>4.2</td>
<td>Percentage of word-final pre-pausal tokens in casual conversation style of /t/ by variant in Tyneside English</td>
<td>115</td>
</tr>
<tr>
<td>4.3</td>
<td>Sound changes in the vowel system of Philadelphia, and which sex leads them</td>
<td>122</td>
</tr>
<tr>
<td>5.1</td>
<td>The eight DJs, their affiliated stations, and their music genres</td>
<td>134</td>
</tr>
<tr>
<td>5.2</td>
<td>The fourteen Likert scales used in the gender perception experiment</td>
<td>144</td>
</tr>
<tr>
<td>5.3</td>
<td>Means and standard deviations for the results of the perception experiment</td>
<td>146</td>
</tr>
<tr>
<td>5.4</td>
<td>Repeated measures one-way ANOVA comparing the Likert scale responses by the 13 participants across the eight DJs</td>
<td>147</td>
</tr>
<tr>
<td>5.5</td>
<td>Loading weights for the three components underlying the perception data</td>
<td>148</td>
</tr>
</tbody>
</table>
Table 5.6 The DJs’ values for the three aggregate scales, *Gender, Social Class, and Personality*…………………………………………………………………………150

Table 5.7 Average zero-crossing rates for the eight DJs …………………155

Table 5.8 Pearson $r$ correlations between zero-crossing rate and the fricative duration for each of the eight DJs……………………………………156

Table 5.9 Token counts and mean zero-crossing rates for two environments, part of a morpheme and not part of a morpheme, for the eight DJs…………………………………………………………………………157

Table 5.10 Token counts and mean zero-crossing rates by sonority of the preceding segment for the eight DJs……………………………………158

Table 5.11 Token counts and mean zero-crossing rates for two environments, before a voiceless segment and before a voiced segment, for the eight DJs………………………………………………………………………158

Table 5.12 Token counts and mean zero-crossing rates by sonority of the following segment for the eight DJs……………………………………159

Table 5.13 Pearson $r$ correlations between zero-crossing rate and the social variables……………………………………………………………………160

Table 5.14 Average zero-crossing rates for each DJ by environment (following voiceless segment or following voiced segment)………………161

Table 5.15 Average zero-crossing rates for each DJ controlling for following environment……………………………………………………………………162

Table 5.16 Pearson $r$ correlations between zero-crossing rate and the social variables, controlling for token counts before following voiced and voiceless segments ……………………………………………………162

Table 5.17 Mean zero-crossing rates and fricative durations before following voiceless and voiced segments………………………………………………163

Table 5.18 Hierarchical multiple regression models for zero-crossing rate before a following voiceless segment and before a following voiced segment………………………………………………………………………164

Table 5.19 The four-way distinction in vowel length, as determined by the vowel’s inherent length and context …………………………………………………167

Table 5.20 Average vowel durations for heavily-accented and lightly-accented vowels for the eight DJs…………………………………………………168

Table 5.21 Contextual and Inherent Vowel Length Contrast Ratios for the eight DJs………………………………………………………………………170

Table 5.22 Pearson $r$ correlations between Contextual Vowel Contrast Ratios and the independent and social variables…………………………173
Table 6.1  List of major and minor dialect regions identified in the *Atlas of North American English* ..........................................................182

Table 6.2  The average distance between /æ/ and /iy/ in mels for each dialect region by sex .................................................................186

Table 6.3  The average distance between /o/ and /uw/ in mels for each dialect region by sex .................................................................191

Table 6.4  List of mergers examined in the *Atlas of North American English*, along with example words elicited .................................198

Table 6.5  Counts and row percentages for three degrees of speaker perception of a merger, “merged,” “close” and “distinct” for men and women for 13 ongoing mergers in North American English .................................................................199

Table 6.6  Counts and row percentages for three degrees of speaker production of a merger, “merged,” “close” and “distinct” for men and women for 13 ongoing mergers in North American English .................................................................200

Table 6.7  The six perception mergers that show significant sex differences........203

Table 6.8  Various link functions for the Ordinal Regression Analysis ............207

Table 6.9  Ordinal Regression Analysis results testing for correlations between FVD, BVD, and the merger categories “merged,” “close” and “distinct” for twelve mergers.........................................................209
List of Appendices

Appendix A: List of Hypotheses ..........................................................................................225
Appendix B: Transcriptions of Speech Samples ...............................................................228
Appendix C: Zero-crossing Rate PRAAT Script...............................................................233
Appendix D: Merger Cross Tabulations..........................................................................235
Introduction: Two Types of Phonetic-level Variation

Natural speech seems to be almost infinitely diverse.
Lindblom et al. 1992:358

1.1 Introduction

Variation at the phonetic level of the speech signal has been of interest to phoneticians for over a century now, and of interest to sociolinguists for several decades. When phoneticians and sociolinguists look at inter-speaker variation, they must consider the following factors, among others:

- sex / gender\(^1\) (variation due to anatomical differences in the vocal tract, variation due to differences in social gender)
- age (variation in child language acquisition, variation due to atrophy of the vocal mechanisms, ongoing language changes)
- context (co-variation)
- speaking style (speech rate, level of formality)

Although both groups of researchers are looking at the same sources of variation in the phonetic signal, their approach is quite different. For example, phonetics research is often done in sound-treated rooms with university students as subjects. Phoneticians use sophisticated mathematical models and equipment such as pneumotachograph masks to record and analyze the phonetic signal, and a wide range of measurements is taken and

\(^1\) Throughout this dissertation, I use the terms “sex” and “biological sex” to refer to the classification of humans into the two categories “male” and “female” based on anatomy. The term “social gender” refers to the gradient social construct that is a product of society. The term “gender” is intentionally meant to be vague, and refers to both the biological categorization and the social construct.
discussed in a single study. Gender is almost always treated as a static dichotic category.

Sociolinguists, on the other hand, use a wide range of subjects from various socio-economic backgrounds and with various education levels, but seldom record their subjects in a sound-treated room. Sophisticated equipment is seldom used as it draws the speaker’s attention to the fact that they are being recorded, and normally only a small number of variables is considered. Sociolinguists are aware that gender is dynamic and can change from utterance to utterance. In short, the emphasis of the phoneticians’ study is on the variation found in the signal itself, whereas sociolinguists tend to focus on external sources of variation. The end result is that phoneticians often struggle with their explanations of the external factors causing variation (we will see an illustrative example shortly), and the sociolinguists tend to look at a limited number of phonetic variables over and over again. There is a clear need for the two sub-disciplines to draw on the strengths of each other. The objective of this dissertation is to reduce the gap between these two approaches in the study of phonetic variation, with a particular emphasis on gender differences.

To do this I begin by presenting a simple classification scheme of phonetic-level variation that emphasizes the distinction between cross-linguistic tendencies that seem to be grounded in the anatomical differences in the vocal tract of men and women, and other variation. I call this classification scheme the “phonetic-level engendered variation framework.” As will be shown in Chapters Two, Three, and Four, making such a distinction is theoretically advantageous in that it leads to a number of testable hypotheses about the roles biological sex and social gender play in language variation and change at the phonetic-level. Several of these hypotheses are then empirically tested with new data in Chapters Five and Six. The most important of these hypotheses concerns the concept of phonetic distinctiveness, a sex-based behavioural difference. Chapter Three elaborates on this concept in detail.

1.2 The phonetic-level engendered variation framework

1.2.1 Two types of phonetic-level variation

The phonetic-level engendered variation framework is shown schematically in Figure 1.1. Phonetic-level inter-speaker variation can be classified as either motivated by sex-based
INTRODUCTION: TWO TYPES OF PHONETIC-LEVEL VARIATION

Engendered variation is phonetic-level variation that is argued to be grounded in the physiological and anatomical differences in the vocal tracts of men and women, and may be either a direct result of the differences, or an indirect result of the differences (i.e., via social stereotypes, see Biemans 2000). As such, engendered variation encompasses variation due to both biological sex differences and social gender differences. Nongendered variation is variation that cannot be consistently explained by anatomical differences between men and women, nor by social gender differences that have their basis in anatomy. Nongendered variation may potentially still have gender-related social meaning associated with it, but there is a key distinction between the two types of variation: cross-linguistic consistency. Engendered variation is expected to show greater cross-linguistic consistency due to its anatomical grounding, whereas nongendered variation is expected to show less cross-linguistic consistency. For example, women on average have higher-pitched voices than men, regardless of language spoken, and therefore average vocal pitch is an example of engendered variation. On the other hand, women do not on average have more nasal voices than men, although there may be specific communities where nasality is associated with female speech. For example, Boas (1911) reports that female speakers of some Inuit dialects use nasals in final position, while males use voiced stops. But the inconsistency across languages means that nasality is not an example of engendered variation.

Each type of variation in Figure 1.1 is further broken down into segmental or suprasegmental. Segmental phonetic-level variation occurs within the temporal boundaries of segments. Supra-segmental variation transcends the temporal boundaries

<table>
<thead>
<tr>
<th></th>
<th><strong>Type I: Engendered</strong></th>
<th><strong>Type II: Nongendered</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmental</td>
<td>Engendered segmental variation (e.g., sibilant centre of gravity)</td>
<td>Nongendered segmental variation (e.g., Canadian Raising)</td>
</tr>
<tr>
<td>Suprasegmental</td>
<td>Engendered suprasegmental variation (e.g., breathiness)</td>
<td>Nongendered suprasegmental variation (e.g., nasality)</td>
</tr>
</tbody>
</table>

Figure 1.1  Two types of phonetic-level variation
of segments. Examples of each subtype follow.

An example of engendered segmental phonetic-level variation is the variation seen in measurements of sibilant centre of gravity. Sibilant centre of gravity is defined as the first statistical moment (i.e., the mean) of the spectrum produced by the frication in the middle of the fricative (Forrest et al. 1988). Sibilant centre of gravity is known to systematically vary by language (Gordon et al. 2002), social class (Stuart-Smith et al. 2003), and sex (Flipsen et al. 1999). The sex differences are due in part to the differences in the size of the resonating cavity anterior to the point of constriction (Flipsen et al. 1999:674; Stevens 2000:398–9). These anatomical differences in the vocal tract of men and women are thought to result in measurable differences in sibilant centres of gravity, making centre of gravity a source of segmental engendered variation. (see §2.3 for further discussion).

An example of engendered suprasegmental phonetic-level variation is the voice quality known as breathiness. Breathy voice quality is defined as inefficient or very lax vocal fold vibration accompanied by audible friction (Laver 1980, 1991). The muscular effort is too low to sustain modal voicing with complete closure of the vocal folds. The incomplete closure during the close phase of the glottal cycle allows a continuous stream of air to escape from the lungs, which in turns gives the perception of vocalizing while breathing. Although breathiness is only relevant to certain segments (specifically, those that involve periodic vibration of the vocal folds in their production), it is, nonetheless, possible to maintain a breathy phonation setting over an entire speech event.

Breathiness has been shown to vary systematically by dialect (Henton and Bladon 1985), politeness level (Ito 2003), and sex (Klatt and Klatt 1990, Södersten and Lindestad 1990, Trittin and Lleó de Santos y 1995). All of the studies of sex-based variation in breathiness show that women produce more breathiness than men, although there is a lot of variability and overlap between the two sexes. In an attempt to explain the apparent consistency of the patterning of the sex, phoneticians have looked for sex differences in vocal fold behaviour. For example, Södersten and Lindestad observed a greater frequency of incomplete closure in the form of a small hole in the centre of the closure, or a glottal chink, in women. A glottal chink during the closed period of vocal fold vibration allows air to escape, resulting in a breathy voice signal (see §2.5). It is not
clear if the tendency that Södersten and Lindestad observed holds cross-linguistically, or even if the sex differences are due to anatomical differences in the vocal folds or just behavioural differences. Regardless of the actual source of the pattern, the pattern seems to be consistent across languages, and that makes breathiness a source of suprasegmental engendered variation.

An example of nongendered segmental phonetic-level variation is “Canadian Raising” (Chambers 1973). Canadian Raising is the process in which the onsets of the diphthongs /aj/ and /aw/ raise to /ʌ/ in certain environments. Canadian Raising is a marker of dialect (Bailey 1982, Chambers 1973), and sex (Dailey-O’Cain 1997). Dailey-O’Cain carried out an impressionistic count of raising on the speech of 30 residents of Ann Arbor, Michigan. The subjects were equally distributed among three age groups and both sexes. Tokens of /au/ were gathered from interviews and word list readings, and then coded as one of four variants, raised, non-raised, fronted, and monophthong. A statistical analysis on the distribution of the raised versus non-raised variants by sex and age (the other two variants almost never occurred) revealed that women raise the onset of /aw/ significantly more frequently than men of their same age group (p.115). There is, however, a priori no difference in the anatomy of the vocal tract between men and women that would suggest that women might raise the onset of the diphthong more often than men. Canadian Raising is therefore a source of segmental nongendered variation.

An example of nongendered suprasegmental phonetic-level variation is nasal voice quality. A speaker with a nasal voice quality tends to pronounce oral segments with the velopharyngeal port ajar, allowing air to leak through the nasal cavity. Seaver et al. (1991) demonstrated that nasal voice quality varies by dialect region and sex. They found that the Mid-Atlantic speakers used a more nasal voice quality than the speakers from the other regions, and that females use a more nasal voice quality than males. They suggest that the sex differences “may reflect subtle differences in underlying anatomical and physiological differences related to velopharyngeal closure for the two genders” (p.720). However, the velum does not display sexual dimorphism (Ohala 1983:fn. 19), and therefore any gender differences are unexpected. Furthermore, a number of aerodynamic studies by Zajac (Zajac 1997, Zajac and Mayo 1996, Zajac et al. 1998) failed to find consistent gender differences in nasal airflow. The lack of physiological differences in
the velopharyngeal port of men and women predicts that men and women will not
demonstrate differences in nasal airflow, and so Zajac’s failure to find consistent gender
differences is not surprising. As such, there is not any reason why we might expect that
one sex would cross-linguistically use a more nasal voice quality or nasalize more often
than the other. This makes nasality a nongendered source of phonetic variation.

In Chapter Two of this dissertation, I present a detailed discussion of five examples
of engendered variation: Voice Onset Time, sibilant centre of gravity, average pitch,
breathiness, and creak. For each of these variables, I present several studies that show a
consistent cross-linguistic sex pattern for that variable, along with a discussion of the
relevant anatomical and physiological differences in the vocal tracts of men and women.

1.2.2 Extending the source of phonetic-level engendered variation

In Chapter Three, I expand the scope of phonetic-level engendered variation to
include sex-based variation that seems to shows a consistent cross-linguistic trend, but is
arguably not grounded in anatomical sex differences in the vocal tract. I label such
variation “behaviourally-determined” engendered variation. One example of
behaviourally-determined engendered variation is sex-based differences in vowel space
dispersion (for example, Henton 1990, Labov 1972, Yang 1996, etc.; see §3.2 for full
discussion). The link between behaviourally-determined engendered variation and
anatomically-determined engendered variation is the apparent cross-linguistic
consistency shown by both. Cross-linguistic consistency is also the characteristic that
distinguishes these two subtypes of engendered variation from nongendered variation.

In Chapter Three I review a variety of studies that examine a range of phonological
contrasts besides vowel contrasts, such as phonemic length contrasts, obstruent voicing
contrasts, vowel reduction, and allophonic contrasts. In every study reviewed, women
produce sharper temporal and / or acoustic correlates of the phonological contrast
examined. I argue this general sex pattern — women produce sharper phonetic-level
distinctions for a given phonological contrast than men — tends to be consistent across
languages, and therefore is another source of phonetic-level engendered variation.

1.3 Prior sociolinguistic studies of phonetic variation

Phonetic-level variation has been examined from the inception of the field of
sociolinguistics. For example, Labov (1966), as part of a sociolinguistic study of the English of New York City, carried out an impressionistic examination of the fronting and tensing of the low front vowel in words such as *bad*, *bag*, *ask*, *pass*, and *dance*. He transcribed each token as one of six variants: \([\text{ɪ}]\), \([\text{ɛ}]\), \([\text{æ}]\), \([\text{æ}]\), \([\text{æ}]\), and \([\text{æ}]\) (p.52). Labov demonstrated that the choice of the pronunciation of /æ/ correlated with speech style (p.114–28), class (p.221), ethnic group (p.295–8), and sex (p.313). Although sex differences were found in the pronunciation of /æ/, when removed from the social background of New York City, they are not *a priori* expected. Thus if we were to repeat the study, in for example, Toronto, Canada or London, England, then we may or may not find similar patterns of variation. This is because we cannot say that a particular variant of the pronunciation of /æ/ such as \([\text{æ}]\), or any other variant for that matter, has an inherently masculine quality or an inherently feminine quality to it.

Today, 40 years later, the emphasis of sociophonetics is still on nongendered segmental variation. Thomas’s (2002) chapter titled “Instrumental Phonetics” in *The Handbook of Language Variation and Change*, which is intended to be a “repository of essential knowledge” (Chambers et al. 2002:2), gives an overview of where the field of sociophonetics stands today. His discussion of recent instrumental studies of variation covers a diverse variety of phonetic-level phenomena:

- /o/ and /au/ fronting
- vowel shifts and other variation
- vowel tense/lax contrasts
- vowel reduction
- gestural sequences (e.g., unrounding then fronting in /oi/)
- diphthong production

All of these studies are of vowels, and Thomas himself points out that “More instrumental studies of consonantal, prosodic, and voice quality variation are needed” (p.190). Furthermore, none of these variables is engendered in that we expect men or women to produce a certain variant more often for anatomical reasons.

It seems, then, that the focus of sociophonetic research is currently stuck in the
upper right-hand corner of Figure 1.1, particularly in North America. The lack of studies other than of nongendered segmental variation is not just an empirical problem, but has theoretical consequences as well. The other types of variation listed in Figure 1.1 other than nongendered segmental variation clearly may be involved in sound change since they vary from one speech community to the next (for example, Esling 2000, Wagner and Braun 2003, etc.), but it is unclear if the generalizations made about sound change and nongendered segmental variation apply to these other types of variation. For example, the distinction between lexical diffusion and gradient sound change may only be relevant to nongendered segmental variation. It is hard to imagine a speaker who adjusts his sibilant centre of gravity or her amount of breathiness depending on the lexical frequency of the word in their daily speech, although, to my knowledge, this claim has never been tested. Yet, although much of sociolinguistic theory is founded on the behaviour of language variation in language change scenarios, the distinction between nongendered segmental variation and the other types has yet to be made on theoretical grounds. Chapter Four presents a more thorough discussion of the theoretical consequences of this distinction between engendered and nongendered phonetic-level variation, where I argue that making this distinction helps us refine and expand our theories of language variation and change.

The field of sociophonetics is not quite as limited in focus as I have made it out to be. There are number of British researchers, notably Stuart-Smith, Foulkes, and Docherty (Docherty and Foulkes 1999, Foulkes and Docherty 2000, Stuart-Smith 1999, Stuart-Smith et al. 2003), who have carried out sociolinguistic research on phonetic-level variation that is not nongendered segmental variation. For example, Docherty and Foulkes (1999) document variation in the preaspiration of voiceless plosives in word-medial position by sex and age for two dialects of British English. They found what appears to be a change in progress, with preaspiration being produced more frequently by the younger generation in Newcastle, but not at all by either generation in Derby. Furthermore, the young women seem to be leading the change, with over 70 percent of their tokens containing preaspiration. In Chapters Two and Four I argue that women should lead this particular change for anatomical reasons, as long as the change remains

---

2 I argue that Voice Onset Time is engendered segmental phonetic-level variation in Chapter Two.
below the level of social awareness.

The major difference between these studies and this dissertation is, although we both discuss engendered phonetic-level variation, these studies are more descriptive in nature or focused on the implications for phonetic theory, whereas mine is more concerned with the implications for sociolinguistic theory. These authors have laid the groundwork by showing us that engendered phonetic-level features index social categories. I now ask: What are the implications of this type of variation for sociolinguistic theory?

1.4 Prior phonetic studies of social variation

To date, there have been a number of phonetic studies of social variation, ranging from Voice Onset Time (VOT) as a marker of African American English (Ryalls et al. 1997), to speech rate as a marker of sex and dialect (Byrd 1994). However, these examples are still the minority, and phoneticians tend to ignore social sources of variation. The lack of serious consideration of social sources of variation in studies of sex-based differences in our speech is can be particularly problematic because, as we shall see shortly, social gender is capable of overriding biological sex. This lack of consideration of social sources of variation is a point that I elaborate on in Chapter Two, so one representative example here will suffice.

Koenig (2000) examined the voicing of /h/ and its relation to the Voice Onset Time of plosives in men and women. Because /h/ in not contrastive for voicing in English, it is possible for a speaker to produce a fully voiced /h/ without a loss of linguistic information, and Koenig refers to several studies that show that the voicing of /h/ occurs intermittently. She took several airflow and acoustic measurements of /p b t d h/ in different environments from seven men, seven women, and seven children. The tokens were extracted from lists of nonsense words. Koenig then used the measurements to make various claims about the mechanical and aerodynamic conditions at the glottis. Of her results, of particular relevance are the percentage of /h/ tokens that are fully voiced for each sex (Table 1.1). Voice Onset Time for /h/ was measured from the zero crossing of the first derivative of the airflow signal to the onset of voicing of the following vowel. A fully voiced /h/ showed unbroken voicing throughout the /h/. Overall, men produced
many more tokens of fully voiced /h/ than women (87 percent versus 35 percent). However, two of the seven women, AF3 and AF6, appear to be outliers, producing 91 percent and 100 percent of their /h/ tokens as voiced. If these two are removed, then the average for the women is only 11 percent. Yet, Koenig fails to draw attention to their anomalous behaviour. This takes on significance when we consider her explanation for the notable difference between the male group and the female group. Koenig points out that the tendency for men to produce more voiced /h/ than women has been observed in other languages such as Japanese, and, therefore, must be a universal tendency grounded in the biomechanics of the larynx. She points out that (drawing on Titze 1989, and others) vocal fold phonation threshold is directly related to:

- coupling stiffness of the vocal folds
- damping of laryngeal tissues
- pre-phonatory glottal half width
- translaryngeal pressure coefficient
- glottal convergence angle

Vocal fold phonation threshold is inversely related to:

- vocal fold thickness
- amount of vibratory phase difference between upper and lower masses

Men have longer, thicker vocal folds with lower stiffnesses than women. Also extra tissue bulk in men may contribute to a smaller convergence angle. The end result is that men have a lower phonation threshold, and therefore may voice more often and sooner than women. This is borne out in the /h/ and plosive VOT data, with men producing more tokens of voiced /h/ and shorter average VOT than women, although the differences in plosives did not reach statistical significance. With regard to the sex differences in VOTh, Koenig concludes that “it therefore seems unlikely that the observed… gender effects on VOTh distributions result from social or other speaker-selected characteristics” (p.1223). But when we consider the two female speakers AF3 and AF6 her argument loses some of its conviction. How could it be that the percent of
voiced /h/ tokens is accounted for by the anatomical differences in the vocal folds of men and women with the social characteristics of the speakers not playing a role, and yet we have two women who clearly pattern with the men? From a statistical perspective, that two of the seven women produce values that are higher than the mean value for the men when there is such a large gap between the means for the men and women is an anomaly, and warrants consideration. Left unexplained, the behaviour of these two women seems to contradict Koenig’s claim that VOTh is unlikely to be a result of the social characteristics speakers does not result from social characteristics.

Koenig’s argument is flawed by her assumption that differences in the speech signals of men and women cannot be over-ridden by social factors. This point is perhaps demonstrated the most convincingly by Andrews and Schmidt (1997) in their study of the voice characteristics of 11 biologically male cross-dressers. Each subject read a short text passage in his natural voice while dressed as a man, and then again on a later date while dressed as a woman. Recordings were presented to undergraduate speech science students who rated each voice sample on 18 perceptual scales without being told that each speaker produced two speech samples. Ten out of 11 subjects were perceived to have a higher pitch during their feminine presentation. Instrumental analysis revealed that average fundamental frequency was significantly higher for the feminine presentation for nine of the 11 men, with one of the men producing an average fundamental frequency during his feminine presentation of 196 Hz, well inside the normal range for women. Men produce a lower pitch than women for anatomical reasons. Yet, this does not mean that fundamental frequency cannot also vary according to “social or other speaker-selected characteristics.”

<table>
<thead>
<tr>
<th>Men</th>
<th>AM1</th>
<th>AM2</th>
<th>AM3</th>
<th>AM4</th>
<th>AM5</th>
<th>AM6</th>
<th>AM7</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent voiced</td>
<td>60%</td>
<td>100%</td>
<td>92%</td>
<td>93%</td>
<td>100%</td>
<td>63%</td>
<td>100%</td>
<td>87%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Women</th>
<th>AF1</th>
<th>AF2</th>
<th>AF3</th>
<th>AF4</th>
<th>AF5</th>
<th>AF6</th>
<th>AF7</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent voiced</td>
<td>25%</td>
<td>0%</td>
<td>91%</td>
<td>17%</td>
<td>13%</td>
<td>100%</td>
<td>0%</td>
<td>35%</td>
</tr>
</tbody>
</table>

Table 1.1 Percentage of fully voiced /h/ by sex. Adopted from Koenig 2000, Table 4
Koenig seems to be treating gender as categorical. A speaker’s biological sex is categorical; a speaker’s social gender is gradient. Although men may be predisposed to voice /h/ more often than women in natural, running speech, there is no reason why a woman could not do the same if for social reasons she chose to, much like the cross-dressers who alter their average fundamental frequency according to whether they are dressed as a man or a woman. Koenig’s failure to recognize this point is not uncommon in the phonetics literature, and illustrates the inadequate treatment of social sources of variation by phoneticians, particularly when that source is not part of the experiment design. We will see more examples in Chapter Two that show phoneticians inadequately dealing with socially-motivated variation.

1.5 The significance of phonetic-level engendered variation

The distinction between engendered and nongendered phonetic-level variation forms the backbone of the framework presented in Figure 1.1. This distinction is important to a study of phonetic-level variation because if a variable is engendered, then the variation shown by men and women is predictable to a certain extent. Furthermore, these sex-based patterns are presumably the same regardless of language or dialect, since sex differences in the vocal tract are universal. The anatomies of the vocal tracts of men and women in Japan differ in the same way as the vocal tracts of men and women in Canada. These characteristics of engendered phonetic-level variation are summarized in the following two points:

- when enough speakers are looked at, sex-based linguistic patterns emerge
- these linguistic patterns seem to be grounded in the anatomical and physiological differences between men and women

There is a casual relationship between these two points: the sex differentiation patterns are consistent from one speech community to the next because the anatomical sex differences are universal.

It may seem from these characteristics that phonetic-level engendered variation is completely determined by the biological sex of the speaker. However, research has shown that the actualization of the sex differences can vary from one community to the
next, a point I discuss in detail in Chapter Two. For example, Henton and Bladon (1985) measure the breathiness of vowels for 61 English speakers from two dialects in England, Modified Northern, and Received Pronunciation. They calculated the noise component of the speech signal for each speaker by comparing the ratio of intensity of the first harmonic to the second harmonic. Because women in general show incomplete closure of the vocal folds more often than men (Södersten and Lindestad 1990), breathiness is a source of engendered phonetic-level variation. Henton and Bladon found that women had much breathier vowels than men, and that their Modified Northern speakers were more breathy than their Received Pronunciation speakers. So we see that the anatomical and physiological differences in the vocal tract only help predict the direction of the differences between men and women. Overlaid on this anatomically-determined trend is the variation due to other social factors, such as dialect, social class, and most importantly, social gender. Furthermore, in Chapter Two we will see examples of reversals, where an individual or a sub-group of the men show the female pattern, and an individual or a sub-group of the women show the male pattern. These points are summarized as follows:

- the extent of the sex differentiation seen in the engendered variation varies from one speech group / community to the next
- there are individuals or groups who for social reasons pattern more like the opposite sex

In Chapter Four, I argue that the distinction between engendered and nongendered phonetic-level variation has significance for understanding of language variation. For example, working-class male speakers attach social significance to sounding masculine (see §4.2.1), and therefore we predict that we will see corresponding phonetic-level engendered variation that correlates with social class. This may manifest itself as the amplification of phonetic-level male trends by working-class men, and the reduction of these trends by middle-class men. The same pattern may also manifest itself among women, assuming that they attach the same social value to sounding masculine.

The significance of the engendered versus nongendered distinction is particularly relevant when it comes to sociolinguistic theories of language change. In the last section
of Chapter Four, I introduce three of Labov’s principles on the role of sex in language variation and change. I argue that one of these principles, namely that women lead linguistic changes that occur below the level of social consciousness, does not necessarily hold at the phonetic level with respect to engendered variation. Even though the variation is below the level of social awareness and therefore void of prestige, it still has what I call “gender social significance.” For example, a breathy voice is most likely associated with femininity, regardless of whether it is involved in a change in progress, or is stigmatized, or is used by a man or a woman, etc. In Chapter Four, I review several counterexamples to Labov’s principle. As we might expect, all of the cases discussed in Chapter Four in which men lead phonetic-level changes that are below the level of social awareness (and therefore counterexamples to Labov’s claim), involve men producing more of phonetic-level engendered variants that are associated with masculinity.

The resolution of these counterexamples within a phonetic-level engendered variation framework accomplishes two things. First, it makes Labov’s principle more robust by eliminating several counterexamples to it. Second, it demonstrates the significance of making the distinction between engendered and nongendered variation at the phonetic level. Once this significance has been demonstrated, then the gap between phonetics and sociolinguistics will become much smaller.
Anatomical Sources of Engendered Variation

Men tend to talk as though they were bigger, and women as though they were smaller, than they actually may be.

Sachs, Lieberman, and Erickson 1973:75

2.1 What is engendered variation?
In the first chapter, I introduced engendered variants as phonetic-level tendencies in the speech signal that on average correlate with anatomical differences in the vocal tracts of men and women. Because anatomical sex differences in the vocal tract are consistent across dialects and languages, the linguistic tendencies are also consistent across dialects and languages. The tendencies are gradient by their phonetic nature, and may be reduced or even reversed, or for that matter, amplified or exaggerated, under the appropriate social circumstances.

We have already seen two examples of reversals: the two women who produced more voiced /h/ tokens than many of the men, and the cross-dresser who produced a high-pitched voice when dressed as a woman. Both cases are clear reversals of the tendencies that men voice /h/ more often than women and that men speak in a lower-pitched voice than women. These cross-linguistic tendencies are founded on well-known differences in the vocal folds of men and women (discussed in detail in §1.4). In the case of the two women who produced more tokens of voiced /h/ than the men, we really do not know if there were or were not mitigating social factors, but in the case of the cross-dressing male, his social gender definitely played a role.

The characteristics of engendered variation introduced in Chapter One are:

- it stems from anatomical differences in the vocal tracts of men and women (e.g., men have thicker vocal folds)
it is reflected in our speech in the form of tendencies (e.g., men tend to voice /h/)

- it is consistent cross-linguistically because anatomical sex differences in the vocal tract are universal

- it is gradient by its phonetic nature

- it is adjusted (amplified, reduced, or even reversed) by social factors

This chapter gives five examples of engendered variation: two at the segmental level, and three at the suprasegmental level. These examples are by no means meant to be exhaustive. Rather, the goal of this chapter is to exemplify each of the characteristics of engendered variation.

### 2.2 Voice Onset Time and Voicing

#### 2.2.1 Variation of VOT by language and dialect

The phonological categories voiced and voiceless for obstruents are most often differentiated by Voice Onset Time. Voice Onset Time (VOT) is defined as “the duration of the time interval by which the onset of the periodic pulsing either precedes or follows release [of the occlusion]” (Lisker and Abramson 1964: 384). There is a vast number of studies of VOT variation by phoneticians. Although these studies have not been carried out in a sociolinguistic framework, they do illustrate that VOT has all of the characteristics of phonetic-level engendered variation.

Consider first that differences in the VOT values for the same phonemic category in different languages vary along a continuum. Cho and Ladefoged (1999) measured and contrasted the VOT for aspirated and unaspirated voiceless stops from 18 languages. They found that languages do not choose the same approximate VOT values for, for example, an aspirated stop. Figure 2.1 shows the mean values for unaspirated and aspirated velar plosives for the languages in their study. What is notable about this figure is that the unaspirated stops and aspirated stops do not form tight clusters at either end of the spectrum of VOT values. Rather, the VOT values make use of the full range. They conclude that “there is a continuum of possible VOTs from which languages may choose” (p.226).

Among these 18 languages we find two, Eastern Aleut (VOT = 75 ms) and
Western Aleut (VOT = 95 ms) that are genetically related to each other, yet show different values for their VOTs of velar stops. This suggests that VOT is susceptible to change as dialects diverge from each other over time. This is further supported by Rosner et al. (2000). They measured the VOT for Castilian Spanish voiced and voiceless stops for 16 male and 16 female subjects aged 19 to 25 years old. When compared to an earlier study on three dialects of Latin Spanish (Williams 1977), they found that Guatemalan VOT values were significantly different from Castilian Spanish regardless of place of articulation. Venezuelan and Peruvian Spanish were also compared but only a subset of the VOT values was different. They conclude that VOT differs by dialect in Spanish.

VOT is gradient in nature and varies from one regional dialect to the next. But can VOT also vary across social class, social gender, or ethnic groups within the same geographical region? Ryalls, Zipprer, and Baldauff (1997) measured the VOT of plosives for 10 Caucasians and 10 African Americans from central Florida, with five men and five women in each ethnic group. They found significant main effects for ethnicity and sex. The African American speakers had a mean negative VOT for the voiced category, but comparable values for the voiceless category. The females had longer VOT means than the males. It appears that these African American speakers use VOT as a marker of

Figure 2.1 Mean VOTs (ms) for velar plosives from 18 languages. From Cho and Ladefoged 1999, Figure 9
ANATOMICAL SOURCES OF ENGENDERED VARIATION

ethnic identity. This demonstrates that VOT can be used as a marker of social group membership.

Now that we have seen that VOT can be a marker of identity, the next question is: Is it possible for bi-dialectal speakers to adjust their VOT when they accommodate or shift from a local dialect to a more standard dialect? As far as I know, this has never been demonstrated. However, Magloire and Green (1999) found a comparable effect for bilinguals. They compared the effect of speaking rate on VOT for English monolinguals, Spanish monolinguals, and English-Spanish bilinguals. They find that speaking rate has a large effect on long-lag English stops (/p/, /t/, /k/) and on negative VOT Spanish stops, but very little effect on the short-lag stops. They suggest a universal constraint on the production of short-lag stops. Relevant to our discussion, they found that the early age bilinguals produced the same effects as the monolinguals, suggesting that they have different targets for each language. If so, then early age bi-dialectal speakers most likely also have different targets for each dialect.

Taken together, these studies provide limited evidence that VOT can and does act as a sociolinguistic marker. VOT targets show subtle variation between languages within the same family and between dialects within the same language. Furthermore, VOT targets also show slight variation between different social groups within the same dialect region, demonstrating that VOT values index social factors.

2.2.2 Variation of VOT and voicing by sex

So far, we have seen evidence that VOT can vary according to social factors such as ethnicity. But what makes it “engendered”? I have already alluded to this in the discussion of Koenig’s work on VOT in Chapter One. Following are several more studies that demonstrate that women produce longer VOTs than men for the aspirates, and that the sex difference may be a product of the anatomy of the vocal tract.

Swartz (1992) measured VOT for voiced and voiceless American English alveolar plosives /t/, /d/ in word-initial position for eight males and eight females. He classified each token as one of six types:

Type 1: plosive release and voicing periodicity begins at the same time (VOT=0)

Type 2: plosive released before voicing (positive VOT)
Type 3: voicing began before and continued up to and through the release (negative VOT)

Type 4: voicing began and stopped before the release, then began again at the point of release

Type 5: voicing began before the release, continued up to it, and then paused, only to resume later

Type 6: there was voicing before and after the release, but ceased some time before the point of release, and started at some point after

Swartz found that Type 2 was the most common for both men and women. Men produced types 1, 3, 4, and 5 more frequently than women, whereas women produced more of types 2 and 6. He also found that women had higher means for both /t/ and /d/. Swartz considered that speech rate might be a factor in these sex differences, and indeed he did find that the men spoke at a faster rate. However, speech rate did not correlate with VOT. He commented that the reason for the differences in VOT between sexes is unknown.

Whiteside and Irving (1997) measured VOT for voiced and voiceless British English stops /p, b, t, d, k, g/ for five male and five female speakers. They found that women had higher mean VOTs for all six stops, although only the differences for /p/ and /d/ were significant. They suggest that the “patterns [seen in the VOT of men and women] could be the result of physical, idiosyncratic, stylistic, and accent-determined factors” (p.462).

Whiteside and Marshall (2001) continue the investigation of sex differences in the production of VOT in British English by examining the acquisition of VOT contrasts children. They were particularly interested in the acquisition of the trend seen in the 1997 study, that women produce longer VOTs. They recorded 30 children, with five boys and five girls in three age groups: 7-year-olds, 9-year-olds, and 11-year-olds, producing /p, b, t, d/ in carrier phrases. They found that on average the 7-year-old girls produced longer VOTs than the 7-year-old boys, but the 9-year-old boys produced longer VOTs than the 9-year-old girls. For the 11-year-olds, the voiceless plosives were longer and the voiced plosives shorter for the girls compared to the boys. Whiteside and Marshall explained
these results by examining the difference between the mean VOT of the voiceless plosive and the mean VOT for the voiced plosive (Figure 2.2). The 11-year-old girls produced a significantly larger VOT category difference in comparison to the 11-year-old boys. They suggest that the greater voiced / voiceless contrast produced by the 11-year-old girls supports the claim of “females using more carefully articulated speech” (p.207). They conclude:

One result of this emerging sex-linked difference was that the girls show a more marked ‘voiced’/‘voiceless’ contrast than the boys. The factors underlying this development can probably be attributed to a number of factors, which include motor speech control, and development, anatomical and physiological sex-linked development differences, and sociophonetic and cultural differences.

There is a second putative explanation for the sex-related differences in VOT seen here: women produce longer aspirated plosives in order to maximize the perceptual distinction between the phonological voiced and voiceless categories. I return to this source of engendered variation Chapter Three.

Scharf and Masur (2002) measured VOT for German voiced and voiceless stops for 40 speakers (ten each of older male, older female, younger male, and younger female). They found that older speakers produced shorter voiced and longer voiceless plosives than younger speakers, and that women produced shorter voiced and longer voiceless plosives than men. In both cases the differences were significant. So again we see that that women are producing a greater voiceless-voiced difference. Their search for an explanation of the sex effects leads them “to the assumption that females might articulate more carefully than male speakers” (p.336).

Allen et al. (2003) examine the amount of inter-speaker variability in VOT production. If social factors such as ethnicity play a role, then we expect that VOT would show at least some inter-speaker variation, depending on how diversified the study population is. In their discussion of previous VOT research, they emphasize that VOT is strongly correlated with speech rate (although the study by Swartz described above did not find a correlation). Allen et al. want to know if speakers’ VOT still differs after controlling for speech rate. They recorded four males and four females reading isolated CVC syllable words from a list, and measured the VOTs for /p, t, k/ in onset position. They used vowel-coda duration as a measure of speech rate, and a hierarchical linear
model algorithm to normalize VOT measures for speech rate. The found that men tended to talk quicker than women, and that speech rate was the strongest predictor of VOT, accounting for 82 percent of the overall variability in the data. The subjects ranked from fastest to slowest were: M3, M2, M4, F1, F2, M1, F4, F3. However, the results also showed that even after normalizing for speech rate, individual talkers still differed significantly from each other, and that speaker identity accounted for 43 percent of the remaining variability (or 8 percent of the overall variability). The ranking of the subjects from shortest to longest average corrected VOT was: F1, M1, M4, M3, F2, F3, F4, M2. Again we see the pattern that as a group women produce longer VOTs than men, although the ordering of the sexes is not significant (Mann-Whitney U test shows that they are not: \( U = 0.773, p > .10 \)). The lack of significance is because there are two subjects pattern more like the opposite sex: F1 produces the shortest average corrected VOT, and M2 produces the longest average corrected VOT. Unfortunately, Allen et al. did not discuss these two subjects. They did, however, conclude that there is variability in the VOT of speakers even after we account for speech rate, and that listeners may use this information to recognize specific voices. If that is the case, then it goes without saying that listeners could also use VOT to identify indexical properties such as gender and social group assuming that VOT varies in a systematic way that correlates with these
categories.

Possibly related to sex differences in voicing contrasts indicated by VOT are sex differences in voicing contrasts indicated by periodic vibration of the vocal folds, as shown by Koenig (2000, discussed in Chapter One). Koenig gathered oral airflow, intraoral pressure, and acoustic signals for /p/, /b/, /t/, /d/, and /h/ produced by seven men, seven women, and seven children in carrier sentences. Women produced longer mean VOTs than men for all four plosives, but the differences were not significant. The aerodynamic measurements were also not significantly different for men and women. As discussed previously, Koenig found a tendency for men to voice /h/ more often than women. She claims that the sex differences in the /h/ data reflect anatomical differences in the vocal folds of men and women (listed in §1.4), and then links the /h/ data with the VOT data by pointing out that if men have a tendency to voice during vocal fold abduction more often than women, we might expect men to show slightly shorter values of VOT in voiceless aspirated plosives (p.1225).

Temple (2000) carried out an impressionistic, categorical study of the voicing contrast in obstruents in Metropolitan French. Temple auditorily classified tokens of phonologically voiced obstruents as either voiced or devoiced / voiceless. The data was read speech from thirty middle-class speakers from two locations at the opposite ends of France, Lille in the northeast, and Bordeaux in the southwest. (Temple does not report the break down of the speakers by sex of region.) Table 2.1 shows the proportion of tokens that are devoiced, broken down by sex and region. The upper table shows the word-initial tokens, while the lower table shows the word-final tokens. For both regions, women produce a greater proportion of devoiced tokens, although the sex effect is negligible in word-final position. These results are consistent with Koenig’s finding that women devoice more often than men.

Fant et al. (1991) summarizes several differences in the speech signal of men and women based on two decades of his research at the KTH Laboratory in Stockholm, Sweden. One sex-based difference they note is that women tend to devoiced vowels more than men in certain contexts, such as when followed by a voiceless obstruent. The researchers recorded eight males and four females reading texts in Swedish. Their report is very terse, as they do not report the vowels measured, the phonological environments,
or their measurement techniques. They found that the female speakers have a greater ratio of devoiced vowels to voiced vowels than the male speakers. The researchers do not report tests of significance on the sex differences.

Considered individually, the results of each study lack generality since the studies focus on specific dialect areas, and use a small number of speakers. Furthermore, the VOT data show a lot of variation both within the individual speaker, and across speaker groups. However, when the results of the studies are considered collectively, two presumably related patterns emerge: women tend to produce longer average VOTs than men, and women tend to devoice more than men, while men voice more than women. The consistency of these patterns from one study to the next suggests that there are real, cross-linguistic differences in the production of VOT and voicing by men and women, the most important characteristic of engendered variation. Of course, we cannot be certain, but it seems that this difference in production is explained by sex differences in the composition of the vocal folds. There were, however, some individual speakers who were exceptions to this trend. Since the studies were not carried out in a sociolinguistic framework, we are not certain if the exceptions are socially motivated. Most likely, the differences in the speech signal that are a result of anatomical differences between men and women are very subtle, but sufficient enough to picked up and amplified by speakers, resulting in social differences playing a much more important role.

2.2.3 Perception of VOT

The last point I want to make about VOT concerns perception. Research in speech perception has shown that the voiced-voiceless distinction signalled by VOT is perceived categorically. For example, Liberman et al. (1961) created synthetic speech stimuli which varied acoustically in equal steps across the voicing continuum. When listeners were presented with pairs of these stimuli in an ABX discrimination test (where A and B are different stimuli, and X is either A or B), they were able to discriminate from stimuli drawn from different phonetic categories (voiced or voiceless), but could not discriminate between stimuli drawn from the same phonetic category, even when the VOT difference between A and B was greater for the within-category stimuli than the between-category stimuli.

This, however, is only half of the story. Later research by Pisoni and Lazarus (1974)
demonstrated that listeners can be trained to perceive within-category differences in VOT. They created a set of stimuli similar to those of Liberman et al., and then trained one group of listeners with the stimuli in random order, and another group of listeners with the stimuli in sequential order. They found that although the ABX discrimination test did not produce significant results, when listeners were presented with a 4IAX discrimination test (one pair, A-A, is always the same, and the other pair, B-X, is always different) those trained with the stimuli in sequential order were able to discriminate significantly above chance, and showed a significant improvement over the other group for the within category stimuli distinctions only. The authors interpret the results as providing evidence for separate auditory and phonetic levels of discrimination, and that listeners can discriminate subtle VOT differences in a continuous mode of perception as well as categorical differences (p.322). If that is the case, then listeners can use VOT as an index of social gender in the continuous mode as well as a marker of phonemic contrast in the categorical mode. Thus, the subtle nature of VOT does not present an obstacle to the claim that VOT is a source of engendered variation.

Table 2.1a Proportions of voiced and devoiced obstruents by region and sex in word-initial position. Based on Temple 2000, Table 2. N equals number of tokens for all speakers in that group

<table>
<thead>
<tr>
<th></th>
<th>Lille</th>
<th></th>
<th>Bordeaux</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>voiced</td>
<td>devoiced</td>
<td>voiced</td>
<td>devoiced</td>
</tr>
<tr>
<td>women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>154</td>
<td>10</td>
<td>105</td>
<td>6</td>
</tr>
<tr>
<td>%</td>
<td>93.9%</td>
<td>6.1%</td>
<td>94.6%</td>
<td>5.4%</td>
</tr>
<tr>
<td>men</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>112</td>
<td>2</td>
<td>86</td>
<td>1</td>
</tr>
<tr>
<td>%</td>
<td>98.2%</td>
<td>1.8%</td>
<td>98.9%</td>
<td>1.1%</td>
</tr>
</tbody>
</table>

Table 2.1b Proportions of voiced and devoiced obstruents by region and sex in word-final position. Based on Temple 2000, Table 3. N equals number of tokens for all speakers in that group
2.3 Sibilant centre of gravity

2.3.1 Centre of gravity studies of /s/

A number of studies on the acoustic characterization of /s/ have shown that women consistently produce frication noise that is higher in frequency than the frication produced by males (see Flipsen et al. 1999 for an overview of the literature). Sibilant frication is produced by forming a narrow constriction between the tongue blade and the roof of the mouth. As the airstream passes through the constriction, it accelerates and hits the incisors downstream from the point of constriction. The impact with the incisors creates the frication noise. The peak frequency in the spectrum is determined by the length of the vocal tract between the point of constriction and the point of impact with the teeth (Shadle 1985, 1991). This is approximately two centimetres for men, and slightly shorter for women, resulting in women having a peak frequency that is approximately 15 to 20 percent higher (Stevens 2000:398–9). Flipsen et al. (1999:674) suggest two possible reasons why the resonating cavity of females’ fricatives is smaller than that of males. The first possibility is that the differences in frication noise reflect anatomical differences in the vocal tract. The head circumference of boys is approximately 0.9 cm larger than that of girls from birth to 18 years of age (Nellhaus 1968:107). These differences in head size presumably underlie the differences in cavity size. However, Strand (1999) is doubtful that the sex differences in the vocal tract are capable of accounting for the full magnitude of the differences seen in the frequency of the frication since “general male-female vocal tract differences … exist mainly behind the area of constriction and the obstacle” (p.88).

This brings us to the second possible reason of Flipsen et al. for differences in the sizes of the resonating cavity: it may be that female speakers are creating a point of constriction that is relatively farther forward than that of the male speakers. They note that statistical analysis of the impressionistic transcription of 3120 tokens of /s/ produced by 12 females and 14 males (120 tokens per speaker) revealed that the female speakers produced significantly more dental tokens whereas the male speakers produced significantly more palatalized tokens. Furthermore, Ladefoged and Maddieson (1996) observe that “in a fricative a variation of one millimetre in the position of the target for the crucial part of the vocal tract makes a great deal of [perceptual] difference” (p.137).
Thus if women desire to produce a higher frequency sibilant, it is simply a matter of producing the point of constriction slightly farther forward in the mouth. The second explanation is not mutually exclusive with the first, and Flipsen et al. concede that it is possible that both structural differences and placement differences are operating concurrently to generate the sex differences. In the rest of this section, we will see evidence that supports this concurrent hypothesis.

Centre of gravity as a way of classifying frication was first proposed by Forrest et al. (1988) in a study of the statistical classification of word-initial voiceless obstruents. They treated the spectrum of the noise produced between the onset of the obstruent and the onset of voicing for the following vowel as a random probability distribution, for which they calculated the first four moments (mean, variance, skewness, kurtosis). These four measurements correspond to the concentration, spread, tilt, and peakedness of the energy distribution across the frequency spectrum. They found that this technique accurately classified voiceless obstruents 92 percent of the time and voiceless fricatives 80 percent of the time. When voiceless sibilants are considered alone, the accuracy increases to 98 percent, indicating that the moments analysis technique is a robust way to describe sibilants. The centre of gravity is another name for the first moment, so named because if the noise distribution in the frequency domain was a solid mass resting on a sharp tip, the centre of gravity is the point at which mass is perfectly balanced.

Researchers have known for awhile that there are perceivable differences between the sibilants produced by men and women. Schwartz (1968) investigated the ability of listeners to identify speaker sex from English fricatives in isolation. He found that listeners could identify speaker sex accurately from [s] and [ʃ] but not from [θ] and [f].

Ingemann (1968) repeated Schwartz’s perception experiment with a larger range of fricatives and found that the accuracy of sex identification corresponded to the portion of the vocal tract in front of the place of articulation, so that back fricatives such as [h] and [x] were more identifiable than [f]. (However, her results show that [ʃ] is the second most identifiable fricative after [h], so clearly it is more complicated than just the distance from the point of constriction to the lips.)

Furthermore, LaRiviere (1974) demonstrated that listeners can identify the sex of the individual speakers better than chance by listening to isolated tokens of the English
fricatives /s/, /f/, /z/, and /v/. Listeners were presented with fricatives in isolation from eight familiar voices. The correct identification rates ranged from 16 percent to 27 percent. The voiced fricatives were better indicators of sex than the voiceless fricatives, and the alveolar fricatives were better than the labio-dental fricatives, with /f/ only marginally above chance. LaRiviere concludes that “there is enough inter-speaker variability in the turbulent portions of fricatives to enable listeners to identify speakers” (p.250).

The above research shows that there is subtle, speaker-specific variation in the production of voiceless sibilants, at least in English. This speaker-specific component allows listeners to judge the sex of the speaker, and seems to depend on vocal tract dimensions. But at the same time there is definitely a social component that does not depend on vocal tract dimensions. One study that demonstrates this social component is by Avery and Liss (1996). The researchers examined what makes a voice sound more- or less-feminine to English-speaking listeners. Thirty-five recordings of male voices were presented to female listeners who rated them for masculinity in a pairwise comparison. Nineteen of the voices were reliably judged as either more-masculine-sounding (MMS) or less-masculine-sounding (LMS). Of these, eight of the voices had similar mean fundamental frequencies, which is believed to be a salient cue to gender (see §2.4). These eight voices were analyzed for intonation patterns, mean formant frequencies, and sibilant moments. Avery and Liss found some significant intonation differences and a significant difference between the F2 of [i], a significant difference in the first and third moments for [s] and the third moment of [ʃ], and more vowel reduction in the MMS voices. Specifically, the LMS voices had higher centre frequencies and negatively skewed energy distributions for [s], and a more diffuse energy distribution for [ʃ].

Quite similar to the previous study is one by Linville (1998). She looked at the phonetic correlates of perceived sexual orientation. Recordings of five English-speaking self-identified gay and four straight men reading a monologue from a play were presented to a group of female listeners who judged each sample for sexual orientation. The listeners correctly identified the sexual orientation of the speaker approximately 80 percent of the time. Each recording was analyzed for the following: [s] duration, [s] peak frequency (the frequency of greatest intensity), modal fundamental frequency, speech
rate, and long term average spectra. Multiple regression analysis showed that the variation in the judgements of sexual orientation were almost entirely accounted for by the variation in [s] peak frequency and [s] duration. That there was a correlation with [s] duration but not with speaking rate indicates that the [s] duration correlation was not an effect of speaking rate.

Taylor (1998) carried out a study of the speech of English-speaking gay men as a marker of membership in the gay community. Taylor constructed a “Gay Community Integration Index,” and then did an impressionistic count of “highly sibilant” /s/ versus other /s/, and affricated /t/ versus other /t/, as well as measuring average pitch for 20 gay men and 10 straight men. Only the impressionistic counts for /s/ type showed a significant correlation with sexual orientation. None of the measures correlated with the Gay Community Integration Index scores.

Another impressionistic study of /s/ is Starks’s (2000) survey of (s) fronting in New Zealand English. She asked 808 New Zealanders four short questions designed to elicit one word responses, and analyzed the response to “What comes before the letter ‘t’?” Each of the /s/ tokens was classified as one of two types: tongue tip lowered or tongue tip raised (also called fronted). When she examined her results by ethnicity and gender, she found that in general male speakers used more non-fronted /s/ whereas female speakers used more fronted /s/. A fronted /s/ is higher frequency than a non-fronted /s/ because the distance between the point of constriction and the obstruction is reduced. It seems that women are amplifying the sex differences with tongue shape. Also Pacific Islanders used more fronted, NZ Europeans used slightly more fronted, Maori used slightly more non-fronted, and Asian used more non-fronted /s/ (Figure 2.3). This study provides further evidence that the articulation of /s/ contains a social component.

Gordon et al. (2002) carried out a cross-linguistic study of fricatives from five Native North American languages, Gaelic and Dravidian. They measure fricative duration, Centres of gravity, and overall spectral shape. The study suffers from impoverished speaker and token numbers (only one token of each fricative type repeated twice, and speaker numbers as few as two). They tested for sex differences, and found that women produce significantly longer and higher frequency sibilants in one language, Chickasaw. It is not clear if the lack of the significant sex differences in the other
languages was a consequence of the small number of tokens and speakers, or if there really is a lack of sex differentiation of sibilants by men and women in these other languages.

A study that demonstrates a clear social component to the production of /s/ is Stuart-Smith et al. (2003). The researchers did an acoustic analysis on /s/ tokens from 31 speakers of Glaswegian reading a word list, divided by age, sex, and class. They found an interesting pattern in the peak frequency data (similar to centre of gravity). Their results are shown in Figure 2.4. The eight groups of speakers are arranged in order from the lowest to highest mean peak frequency of /s/. The females, who have higher frequency frication noise for /s/, group together on the right, and the males, who have lower frequency noise than the women, group together on the left. There is a noticeable gap between the males and the females. There is, however, one group that stands out as exceptional: the working-class girls group more with the males than with the females. The researchers explain results as follows (2003:1853–4):

The consistent finding of sex suggests aspects of the acoustic energy of /s/ in male and female speakers are determined by sex, i.e. by anatomical differences leading to resonance differences... Thus, aspects of the acoustical energy indexical of biological or anatomical sex provide a frame within which other aspects of acoustic energy, indexical of gender, may be manipulated.
In general, anatomical differences between men and women either directly or indirectly determine the direction of the patterning of sex, viz., women produce higher frequency sibilants while males produce lower frequency sibilants. Overlaid on top of this trend are the social differences due to age, class, and social gender. The researchers conclude that the working-class girls produce sibilants with low frequencies for social reasons.

Another cross-linguistic comparison is that of Heffernan (2004). He compared the first four moments for tokens of /s/ from 22 speakers, ten Japanese (five female and five male), and 12 English (six female and six male) taken from word list readings. The first moment, centre of gravity, showed a significant sex effect. The sex difference for both language groups, as shown in Figure 2.5, is much smaller than the 15 to 20 percent difference claimed by Stevens (English 10.5 percent; Japanese 5.1 percent). Harmonic to Noise Ratio (HNR. see §2.5) was also measured for each speaker in order to test the hypothesis that sibilant centre of gravity and breathy voice quality (as measured by HNR) are working together to construct the social gender of the speakers. The Pearson’s correlation coefficient for the centre of gravity values and the HNR values was significant ($r(4) = -0.90$, $p = 0.018$) for the English women only. The smaller gap between centre of gravity values for the Japanese men and the women in Figure 2.5, and the lack of a significant correlation led Heffernan to conclude that sibilant centre of gravity plays a more central role in the construction of social gender for the English group than for the Japanese group.

Figure 2.4 /s/ peak frequency for eight groups of Glaswegians. Adopted from Stuart-Smith et al. 2003, Figure 1. Data approximated from their figure.
In every study of sibilants discussed here, and as I mentioned in the introduction to this section, any significant differences reported were always the same: women as a group have a higher centre of gravity than men. The consistency of this result across languages suggests that there is a universal, anatomical component to the sex differences: the distance between the point of constriction and the front teeth is smaller for women. At the same time, many of the studies introduced here show clear evidence for a socially acquired component. It seems that Flipsen et al. were correct when they speculated that the production of /s/ contains both an anatomical and a social component.

2.3.2 Centre of gravity studies of /ʃ/

A relevant question that was not answered in the previous discussion is: Can the other voiceless fricatives also be sociolinguistic markers? The perception experiments such as those by Schwartz (1968) and LaRiviere (1974) suggest that only the sibilants contain enough speaker-dependant information to do so. But it may be that /ʃ/ is also limited in its ability to index social information.

Consider the results of Nittrouer et al. (1989). They looked at acquisition of the phoneme /s/ by English-speaking children. They argue that children start out with the
word or the syllable as the basic unit of speech, and then over time reorganize their speech cognition so that the phoneme is the most basic unit. To test their hypothesis, they look at fricative-vowel utterances of children of different ages and adults. They found two trends: as children get older, the extent of the differentiation between /s/ and /ʃ/ increases, and the extent of the coarticulation of the fricative with the following vowel decreases. Of interest is the /s/-/ʃ/ contrast made by the children compared to the adults. Figure 2.6 shows the average centres of gravity for /s/ and /ʃ/ at 100 ms before the onset of voicing of the following vowel by age. Since children have small vocal tracts, we expect that they should produce sibilants with higher frequency frication than adults who have larger vocal tracts, and that as the children grow older the frequency of the frication noise should lower. This is what happens for /ʃ/ in Figure 2.6, and although we see this trend in the children for the centre of gravity of /s/, there is a dramatic adjustment in the /s/ produced by the adult women. This results in an average adult /ʃ/ that is lower frequency than that of the children, but an average adult /s/ that is higher frequency than that of the children. Women must be adjusting their /s/ to match their social gender by increasing the frequency band of the frication noise, and thus exaggerating their physiological diminutivity.

Nittrouer et al. also examined the frequency at which the formant-like peaks in the noise 30 ms before the onset of vowel voicing occurred, and found that the ratio of /s/-
formants to /ʃ/-formants is approximately the same for children and adults. This ratio is determined by the relative lengths of the back cavities for /s/ and /ʃ/, and therefore the relative placement of the tongue tip. The constant ratio across age groups implies that the relative placement of the tongue is approximately constant. Based on this, they conclude that any differences in the centre of gravity ratios as a function of age must be due to differences in tongue shape (p.130). Why would adults, particularly women, alter their tongue shape to create a higher frequency /s/? There are two reasons. The first reason is to express social gender, as discussed in the previous section. The second reason is something that has not been mentioned in the sibilant literature, but was a common theme in the discussion of VOT — to increase the perceptual contrast between the alveolar and post-alveolar fricative categories. As expected, women in this study produce both /s/ and /ʃ/ with mean centres of gravity that are higher frequencies than those of men. Furthermore, the ratio of /s/ centre of gravity to /ʃ/ centre of gravity is largest for women, followed by men, and then children. We can pursue the perceptual contrast enhancement argument one step further by hypothesizing that the greater distinction produced by women is easier to perceive. If so, then women speak more clearly (at least as far as the perception of these two phonemes is concerned), a conclusion that is familiar from the VOT discussion.

In the previous section, we concluded that the production of /s/ was primarily social, but that the pattern was likely determined by anatomical differences. In contrast, it seems that, at least from the one study presented here, that the social component of /ʃ/ is much smaller than that of /s/. However, more sociolinguistic studies of /ʃ/ are needed before we can truly draw any conclusions.

In summary, the centre of gravity for /s/ seems to contain an anatomical component that results in the cross-linguistic trend of women producing higher average centre of gravity frequencies than men. The average centre of gravity frequency values vary in a gradient manner from one speaker to the next, and from one speaker group to the next. Similar to the VOT data, there are also clear exceptions to the trend. Unlike the VOT studies, at least one of the sibilant studies, Stuart-Smith et al. (2003), was carried out in a sociolinguistic framework. This study demonstrated that social factors play a major role in the determination of the frequency characteristics of /s/. We can conclude that sibilant
centre of gravity is an excellent example of phonetic-level engendered variation.

### 2.4 Pitch

#### 2.4.1 Pitch as a marker of gender

Fundamental frequency (F0) differences between men and women are directly related to the physiological differences of the larynx. These differences include vocal fold length, vocal fold thickness, and the elastic properties of the tissue, of which vocal fold length is the primary factor (Titze 1989). The male vocal folds are approximately 1.6 times longer than the female’s. This results in a higher average F0 for women than men (see Baken and Orlikoff 2000:Table 6-2 for a summary of 17 studies of F0).

Fundamental frequency, and its perceptual correlate, pitch, is perhaps the most salient cue to the sex of the speaker. Lass et al. (1976) recorded speakers sustaining six different vowels with both modal and whispered voice. The modal voice recordings were then 255 Hz low-pass filtered to produce three sets of stimuli: modal, low-pass filtered, and whispered. Listeners guessed the sex of the speaker for each vowel. Sex identification was 96 percent accurate for the voiced tape, 91 percent accurate for the filtered tape, and 75 percent accurate for the whispered tape. The investigators concluded from this that F0 is more important a cue to the identification of speaker sex than vowel formants.

Research such as that by Lass et al. demonstrates that F0 may index categorical sex differences. But what about gradient gender differences? Does our average pitch reflect our social gender. One might think that the answer would be an emphatic yes, but the evidence is for this is ambiguous, as shown by the following studies.

Studies of stereotypes, such as Valentine and Damian (1988), suggest that F0 plays a role in the construction of social gender. They investigated stereotypes of the ideal voice in Mexico and the United States. They found that university students in the United States feel that the ideal male voice is low in pitch and medium in volume, whereas the Mexican students feel that the ideal male voice is medium in pitch and loud in volume. Both groups of students feel that the ideal female voice is high in pitch and soft in volume. The differences in the perception of the ideal male voice are clearly social in origin. If there are discernible differences in the production of average F0, then we can
conclude that F0 does play a role in the construction of social gender.

In order to sort out how much of our average F0 is anatomically determined, Graddol and Swann (1983) measured the average F0 for 15 male and 15 female English-speaking university students producing a sustained /a/ vowel at the lowest pitch possible (basal F0), the reading of a “boring” passage about interior lighting, and the reading of an “interesting” conversational passage. Correlation of the average F0 for the three speech styles with speakers’ heights and weights produced significant results for the men only, and only between height and average F0 in the boring passage and the basal F0. Graddol and Swann suggest that this shows that as the speech becomes more natural, it moves away from the natural average fundamental frequency determined by physiological factors. The fact that women showed no correlation implies that either their body measurements are not a good indicator of their natural F0, or they avoid it for social reasons. Correlations of basal F0 with the reading passage F0 were significant for men only. The authors argue that if we assume that basal F0 is due to physical factors, and since average F0 and basal F0 do not correlate with each other for the women, then average F0 is not determined by physical factors for the women. If average F0 was physically determined for women, then their basal F0 should correlate with average F0. The researchers conclude that for social reasons men adopt a speaking fundamental frequency that reflects their body build, but women do not (p.363).

A number of phoneticians, motivated by the belief that gay men sound more effeminate than straight men (Lerman and Damsté 1969:341), turned to the investigation of the voices and the stereotypes surrounding gay men and women. The following studies use different elicitation techniques and sample sizes, but they have one thing in common: they examine average F0. It was perhaps assumed that pitch, being such a salient marker of sex, would also be a clear marker of sexual orientation (and therefore social gender).

Lerman and Damsté (1969) compared the mean F0 of 13 self-identified gay men with the mean F0 of 13 straight men. Each subject was recorded reading a passage, describing a picture, and producing a sustained /a/ and /i/. No significant differences were found for average F0 measurements between the two groups of men for any of the speech samples. The authors suggested that the lack of significant findings may be due to either the fact that no real differences exist, or to the fact that gay men only use a high-
pitched voice in their own “social milieu” (p.344).

Guadio (1994) examined the stereotype that gay men have greater pitch range / more dynamic pitch than straight men. The subjects, four self-identified gay men and four straight men read two texts, one an accounting passage, the other a monologue from a play. From these, 15-second excerpts were taken. These samples were presented to listeners who graded them on four scales (straight–gay, effeminate–masculine, reserved–emotional, and affected–ordinary). As might be expected, all of these scales correlated with each other, with the exception of reserved–emotional. An acoustic analysis was also carried out on each speech sample to determine if several measures of pitch range and pitch dynamism correlated with the perceptual scales. No significant results were found. Guadio concluded “pitch range and pitch variation do not by themselves crucially affect whether or not a man will be perceived as ‘sounding gay.’” (p.53).

Linville (1998, described in §2.3.1) recorded five self-identified gay and four straight men reading a monologue from a play. Measurements of mean F0 did not correlate significantly with sexual orientation.

Smyth et al. (2003) collected voice samples from eight straight and 13 self-identified gay men reading a scientific passage, reading a dramatic passage, and responding to an open-ended question that was designed to elicit a more spontaneous speech sample. The speech samples were then presented to two groups of listeners, one consisting only of gay men, and the other of university students, who judged whether the voices sounded gay or straight. The speech samples were also analyzed for average F0. The relationship between “straight / gay” ratings and mean F0 was not significant.

The lack of significant results between the correlation of mean pitch and sexual orientation has lead some researchers to claim that “mean fundamental frequency alone is not a cue to sexual orientation” (Rogers and Smyth 2002). In other words, it appears that based on these experiments that F0 is not a salient index of social gender. However, as I said, the problem is not so straightforward. For one thing, when men transfer from one sex category to the other (i.e., they consider themselves to be women), then the role of F0 is much clearer.

Bralley et al. (1978) took F0 measurements and gender perceptual judgments over
seven voice therapy sessions for a male-to-female transsexual. They show that at the end of the sessions, the subject had a higher average F0, a larger pitch range (maximum F0 minus minimum F0), and was perceived as sounding more feminine.

Wolfe et al. (1990) also looked at the F0 and intonation for 20 male-to-female transsexuals. They were recorded during conversation, and then the speech samples (about 10 seconds) were presented to two different groups of listeners: one group rating them as either male or female, and the other group rating them on a gender scale. Nine were rated as women, and 11 as men. Acoustic measures showed that F0, the extent of downward intonation (without interruption), the percent of level intonations, and the percent of level shifts (intonation pattern is held the same across an interruption) correlated significantly with the sex judgments. The gender ratings also correlated significantly with average F0, percent level shifts, and the extent of downward intonations. The results show that the feminine voices are perceived as more variable in their intonations and shifts, with exception of the downward shifts, of which the male voices showed the greater range. The authors suggested that this is because the female voice subjects are avoiding the lower frequency range. They also point out that an examination of individual voices shows that average F0 is much more important than intonation in making the sex judgements, and that the voices with an average F0 below 150 Hz tended to be judged as male while those voices with an average F0 above 160 Hz tended to be judged as female.

The strongest evidence for the indexation of gender by average F0 is provided by Andrews and Schmidt (1997). This experiment examined the voices of eleven biologically male cross-dressers. Each subject read a text passage dressed as a man, and then returned a day later and reread the same passage dressed as a woman. The recordings were presented to undergraduate speech science students trained in the perception of voice. Each listener rated each voice sample on 18 perceptual scales (such as strong–weak, clear–full, slow rate–fast rate, etc.). Paired t-test comparisons between the masculine and feminine presentations were carried out on the listeners’ judgements. Every subject was found to be significantly different on at least four scales (including feminine–masculine), and some on as many as 15 of the 18 scales. Ten out of eleven subjects were perceived as using a higher-pitched voice during their feminine
presentation. At least eight of the eleven feminine presentations were also considered more breathy and more animated. Besides the perception test, each recording was also measured for average F0, F0 range, entire duration, and duration of stressed vowels. Mean duration for stress vowels and F0 range measurements were not significant. Entire duration was significant, with the feminine presentations taking longer. The mean F0 was significantly higher for nine of the eleven feminine presentations. The authors suggest that a speaker attempting to project a feminine image by using a higher pitch voice may sound less credible when attempting upward pitch inflections, and therefore some of the speakers may avoid it. They point out that the speakers with the low F0 in their feminine presentations were quite variable in the intonation, and this may have helped them be perceived as feminine.

So it seems that some of the studies of pitch found significant correlations, while others did not. These contradictory results can be made sense of if we assume that that average F0 is used as a cue to sex, and not social gender. Gay men may be gay, but they are still men, and their mean F0 marks them as such. Cross-gendered men on the other hand consider themselves to be women, and their mean F0 tends to mark them as such. Some of these women are better at overcoming the predisposition their male vocal fold anatomy has for a low-pitched voice than others. Either way, F0 is robust marker of sex for the transgendered men.

But this conclusion, that F0 tends to index sex but not gender, is certainly only half the story. The lack of sociolinguistic sophistication of the phonetics research may also play a role in the lack of significant findings for the studies of pitch in gay-sounding men. For example, Mendoza-Denton and Jannedy (1998) report in their ethnographic study of California Latina gang girls, “one speaker that we have analyzed speaks in a very high-pitched, stereotypically interpreted as feminine voice when talking with her boyfriend or with her mother, and a low-pitched creaky voice when talking with her girl-friends or when she is telling a fight story.” Lerman and Damsté were probably on the right track when they claimed that those gay men who do use high pitch use it in their own social milieu. It may be that if phoneticians were to utilize the more sophisticated ethnographic techniques such as those used by Mendoza-Denton and Jannedy, then they may see significant results. Regardless, it is clear from this discussion that fundamental frequency
is a source of engendered phonetic-level variation. For anatomical reasons in general males tend speak with lower-pitched voices than females. But these tendencies can be manipulated for social purposes on an individual by individual basis. This claim is further supported by the studies of average vocal pitch as a marker of other social categories, such as social class, discussed in the next section.

2.4.2 Pitch as a marker of other social categories

In the previous section, we saw that the average pitch of a person’s voice is a good indicator of their sex, both biological and psychological. We also saw one tentative example that pitch indexes social gender, along with a number of studies that failed to find such an indexation of gender. With only one positive example, it is difficult to claim that pitch acts a sociolinguistic marker. In this section, we will see more evidence that supports this claim, including claims that average pitch is undergoing a change in progress in some English speaking areas and that average pitch is used to mark ethnic group membership.

First consider the evidence for change in progress. Majewski et al. (1972) compared the F0 of 103 Polish male university students with 157 American male university students. They found that the average for the Poles was 137.6 Hz while the average for the Americans was 129.3 Hz, a difference that is statistically significant. They also measured weight and height of the students, but did not find a significant correlation with F0. They surveyed a number of older studies and pointed out that there is a trend that average F0 is lowering in the United States. The lack of a significant correlation with the body measurements suggests that the change of F0 over time in the United States is not due to changes in the physical make-up of American men.

De Pinto and Hollien (1982) measure the F0 of recordings of Australian women made in 1945, and compare the results with modern-day Australian women of the same age. They conclude that for reasons yet unknown, women 40 years ago had a higher average F0.

Hollien et al. (1997) tested three hypotheses on F0 in American English: that F0 is lowering over time; that F0 varies by social class; and that F0 varies by speech material. They did an exhaustive literature review, including a plot of the trend over a 50 year
period. It appears that F0 is gradually dropping for both males and females. They chose three earlier studies for statistical comparison with their own study, one of females, and two of males. All of the studies used between 18 and 25 subjects, the same reading material (Rainbow Passage), and the same measurement techniques. The results from their own study in comparison to the other studies supports a trend of lower F0 over time for the men only. Next, they report data gathered in 1973 from 157 male university students, and 142 males drawn from the enlisted ranks of a nearby Naval Base (Table 2.2). They also gathered height and weight data. They found that the military men used a significantly lower F0, even though the university students were slightly taller and slightly heavier. They point out that the F0 pattern is the reverse of what is predicted from the physical data. They go on to say that the only factor which might account for the results is the slightly greater number of smokers in the military population, but they are not confident since the subjects are young, and so could not have been smoking for very long. They report on the reading versus talking data from the 1973 study, along with an additional 25 males and 18 females. The F0 was significantly lower for the spoken sample. They point out that one in five subjects showed a reverse trend, but explain that this may have been due to the subject getting excited while talking. I suspect that social class is a better explanation than smoking of the lower F0 among the military personal, since working-class speech has a closer association with masculinity than middle-class speech (see §4.2.1). Either way, this paper clearly demonstrates that F0 is a sociolinguistic variable that correlates with both group identity and speech style.

Studies that examine pitch characteristics in the speech of African Americans and Hispanic Americans are another source of evidence that pitch acts as a marker of group identity. Hudson and Holbrook (1981) recorded 100 female and 100 male African American university students reading the Rainbow Passage, and took weight and height measurements for each subject. They compared the average F0 of these students with the results from Fitch and Holbrook (1970), who used the same techniques for recording and analyzing the same number of Caucasian American students. They found that the African American students had a lower mean F0 and a greater mean range. The Caucasian American subjects showed a greater range below the mean mode, whereas the African

---

3 See, for example, Gilbert and Weismer (1974) for the link between fundamental frequency and smoking.
American students showed a greater mean range above the mean mode. Although Hudson and Holbrook did not test for significance, the differences between the two ethnic groups are highly significant (two-sample t test, d.f. = 99, p < 0.0005 for both the men and the women). They did not find significant correlations with the physical measurements. They conclude that if additional studies substantiate the F0 differences between the races, then “structural differences explaining them should be explored” (p.199). These researchers do not acknowledge the possibility of a social explanation.

Morris (1997) looks at the F0 of 45 Caucasian American and 45 African American boys from three age groups (8, 9, and 10 year olds), from two different communities. The boys read a text passage and described a picture. The researchers calculated modal measures instead of means. No significant effect was found for age, speech style, or race. They also measured the standard deviation for each boy. Here they did find results: as the African American boys grow older, the standard deviation increases, but the Caucasian Americans do not exhibit this pattern. They suggest that this is a gradual increase in the approximation of the adult speech model (African American adult males have a greater sigma than Caucasian American adult males).

Awan and Mueller (1996) recorded 35 Caucasian American, 25 African American,
ANATOMICAL SOURCES OF ENGENDERED VARIATION

and 35 Hispanic American 5-year-old children using a picture description technique and simple question and answers. They also measured standing height and weight. The speech samples were analyzed for mean F0, minimum F0, maximum F0, pitch sigma, and speaking range. The results show that the Hispanic group had the highest average F0, and the African American group the lowest, with the Caucasian Americans in between. The difference in the mean F0 of the African American and the Hispanic American boys was significant. They also found that the Hispanic children have a significantly smaller speaking range when compared to the other two groups. The lower average F0 of the African American children agrees with the results of Hudson and Holbrook, but not of Morris. The researchers, however, instead put forth a “speech style” explanation rather than an “ethnic group” explanation. Following Kayser (1983), they claim that Hispanic children “often show different styles of interaction with teachers and adults than they would with other children” (p.577). The implication here is that the Hispanic children are speaking in a more reserved, formal manner in front of the investigator, who is a Caucasian American female. This in turn implies that the African American children, who produced the lowest average F0, are completely at home in front of the investigator, or at least more so than the Hispanic children, and that the differences between the groups are accounted for by differences in the style of speech used during the interview. According to the researchers, further evidence for this claim is provided from the average speaking range of the Hispanic children, which was about a semitone narrower than the average speaking ranges of the other two groups of children. Awan and Mueller suggest that the Hispanic children may behave differently around the interviewer because their home language is Spanish. They say, “some bilingual children have been observed to experience conflicts between their home language and the dominant culture that may be reflected in terms of withdrawal” (p.577). In other words, Awan and Mueller are equating a reduced speaking range with withdrawal.

At first glance, this explanation seems improbable. However, the findings of F0 differences in 5-year-old children is startling considering that children of the same age, when compared within the ethic group and across sexes, do not show significant sex differences in average F0 (Bennett 1981, Bennett and Weinberg 1979a, Bennett and Weinberg 1979b, Busby and Plant 1995, Perry et al. 2001, Sachs 1975, Sachs et al. 1973). These researchers conclude that children do not alter their F0 to approximate the F0 of
the adults of the same sex. Considering this, then it seems unlikely that the children in Awan and Mueller’s study are altering their F0s to closer approximate adults of the same ethnic group. In light of this, a style explanation seems more plausible. Either way, the important point is that there are average F0 differences between the groups of boys.

Hudson and Holbrook (1982) is a continuation of Hudson and Holbrook 1981, but this time looking at reading versus natural speech data with the same speakers. They found that for the natural conversational speech data, mean modal values were significantly lower than reading speech data, mean total vocal range was significantly greater, and mean vocal range below and above the mode was significantly greater. The men had a larger range than the women. The differences between the natural speech data and the read speech data were comparable across sex. A comparison with data for Caucasian American students gathered in a different study show that the Caucasian American students are less variable (smaller sigma).

Fitch (1990), on the other hand, found that F0 is significantly higher when measured from sustained vowels, but did not find a significant difference between reading and connected speech.

2.4.3 Pitch as a marker of politeness

Another source of variation in average fundamental frequency is the politeness level of the speech. Loveday (1981) looked at the pitch levels of men and women in Japanese and English. He had two females and three males from each language read out a dialogue. He found that Japanese women used an artificially high pitch such as 450 Hz when reading the expressions of politeness, but Japanese men did not, preferring instead to use a low-pitched voice. This resulted in a large average frequency difference between the Japanese women’s expressions of politeness and the Japanese men’s expressions of politeness. Both the English speaking men and the English speaking women somewhat raised their pitch during the politeness expressions, with the men doing more so than the women. This resulted in a very narrow average frequency difference between the English speaking men and the English speaking women.

Yamazawa and Hollien (1992) looked at the F0 of Japanese women compared to American women. They recorded 32 Japanese reading a Japanese passage, and for those
that were capable, reading the Rainbow Passage in English. They also recorded 24 Americans reading the Rainbow Passage, and if capable, reading the Japanese passage. For each recording, they measured the average F0 and constructed a histogram of F0 distribution for 4,210 wave pulses. They found that the Japanese tended to produce a bimodal distribution, whereas the Americans tended to produce a unimodal distribution. The bimodal distribution is explained by the H-L tone contrast in Japanese. Furthermore, the bilingual speakers tended to produce the same distribution patterns as their first language even when reading in their second language, showing a heavy first language influence. Overall, the Japanese exhibited a significantly higher mean F0 for all experimental conditions. The difference was greatest when speakers read in their first language, but still existed when they read in their second language. This shows that Japanese women tend to read with a higher F0 than American women. The authors conclude that “while there may be a racial component contributing to this difference [in F0], it probably results primarily from differences [in the use of pitch] between the two languages” (p.139).

Van Bezooijen (1995) built on previous work by Loveday (1981) and Ohala (1983, 1984) by looking at the social motivation for the reported differences in pitch between Japanese and western women. She poses three hypotheses: 1) pitch has social meaning in both Dutch and Japanese culture; 2) there is a strong differentiation between the ideal man and the ideal woman in Japan, and that the woman is characterized by powerlessness, whereas in the Netherlands the differentiation between the ideal man and the ideal woman is much weaker; 3) since high pitch correlates with physical diminutivity, high pitch in Japanese woman will be evaluated positively, but medium or low pitch will be evaluated positively in Dutch woman.

She recorded Japanese and Dutch women reading a text passage in their native language, and then manipulated the pitch to create a low pitch, unaltered, and high pitch sample for each speaker. She played back the samples to both Japanese and Dutch listeners of both sexes. Each listener judged the samples on a seven-point scale for the attributes short-tall, dependent-independent, and modest-arrogant. These judgements correlated with pitch, but there was no interaction with culture, supporting hypothesis one. Each listener also judged each sample for the attribute attractive-unattractive.
Japanese listeners rated the high-pitched voice as the most attractive, while Dutch listeners rated the unaltered voice as the most attractive, supporting hypothesis three. Finally, each listener also imagined and rated an “ideal” male and female speaker of their native language. There was a much larger gap between the average ratings for the ideal Japanese male and female in comparison to the ideal Dutch male and female. The Japanese results straddle the Dutch results so that the ideal Japanese male is more strong, taller, more arrogant, and more independent than the ideal Dutch male, and the ideal Japanese female is less of all of these than the ideal Dutch female. This supports the third hypothesis. The results of this experiment suggest that the ideal female Japanese voice is high in pitch since this portrays an image of petiteness, whereas the ideal female Dutch voice is lower in pitch.

The studies of average fundamental frequency discussed in this section show that women as a group consistently produce a higher average fundamental frequency than men. The explanation for this sex difference is found in differences in the average length of the vocal folds. Much like VOT and sibilant centre of gravity, this cross-linguistic trend in pitch sex differences displays socially motivated adjustments and reversals at the level of the individual. The anatomically-grounded trend, along with the exceptions to that trend show that pitch is a source of engendered phonetic-level variation.

2.5 Breathiness

2.5.1 Breathiness as phonetic-level engendered variation

In Chapter One, I reported that breathy voice quality is result of low muscular effort vocal fold vibration, causing audible frication. This occurs when the vocal folds do not completely close during the close phase, allowing air to continuously escape from the lungs. There is nothing in this definition that suggests that one sex might use breathy voice quality more often than the other sex. However, phonetic studies of breathiness have shown that women use breathy voice quality more often than men, and that the sex difference is a result of differences in the way the vocal folds of men and women vibrate.

Södersten and Lindestad (1990) used a fiberscope to directly observe the vocal fold behaviour of men and women during speech production. The authors videotaped the vocal folds of nine women and nine men producing a sustained /i/ in modal voice at three
pitch levels and three loudness levels for each pitch level. The videos were visually examined and each subject’s vocal folds were rated for degree of incomplete closure. As well, each subject was rated impressionistically on a breathiness scale by a panel of voice experts. In general, pitch was not a significant factor in incomplete closure, while loudness was; the softer the voice, the more incomplete the closure of the vocal folds for both sexes. Overall, women had greater degrees and more instances of incomplete closure than men. Women were also judged to be more breathy. A majority of women had at least some degree of incomplete closure, showing that it is the norm for them, whereas the majority of the men had complete closure. Södersten and Lindestad are not sure why female vocal folds tend to phonate with incomplete closure. They state that “one reason could be cultural factors that might influence vocal behaviour differently in women and men” (p.610).

Klatt and Klatt (1990) examined the production, perception and synthesis of breathy voice quality. They recorded ten women and six men reading two sentences. They determined the ratio of the first harmonic amplitude to the second harmonic amplitude, and the periodicity of the signal in the spectral region of the third formant. Breathy speech samples are known to have a larger ratio than modal speech samples (Huffman 1987), and more noise in the high frequency region of the signal. They found that the ratio of first harmonic amplitude to second harmonic amplitude was lower for men than women, indicating that the females were more breathy than their male counterparts, but there was a large amount of intra-sex variability. The also found that breathiness increases for unstressed syllables, utterance-final syllables, and in the environment of voiceless consonants.

Trittin and Lleó (1995) repeated the Klatt and Klatt experiment for Spanish speakers. Trittin and Lleó also found that on average the ratio of first harmonic amplitude to second harmonic amplitude was lower for men than women. However, their results differed from the results of Klatt and Klatt in that the difference between the two sex groups was smaller for the Spanish speakers than that for the American speakers in the Klatt and Klatt study, with the Spanish females being notably less breathy than the American females. Trittin and Lleó conclude that the female Spanish speakers are only slightly more breathy than the male Spanish speakers, and that the sex difference for the
Spanish speakers is much less than the sex difference found by Klatt and Klatt for the American speakers.

Wagner and Braun (2003) measured fundamental frequency, harmonics-to-noise ratio (HNR), jitter, and shimmer for 50 young adult male speakers from three languages: Italian, Polish, and German. HNR, the ratio of the periodic energy to the aperiodic energy, is another measure of how noisy, and by inference breathy, the speech signal is (for example, Ferrand 2000). Wagner and Braun found significant differences between the groups of men for all of the measures other than jitter. They conclude that there are voice quality differences between the languages. These differences include the amount of noise in the speech signal, although it is not clear if the perceptual correlate for this noise is “breathiness,” or as Wagner and Braun describe it, “roughness.”

I previously reported sex-based voicing differences found by Fant et al. (1991). They also note that their female subjects show a greater tendency for breathiness (p.524). They presents spectrograms from four speakers of Swedish, two men and two women, showing that women tend to produce a breathy vowel before stressed aspirated stops. Fant et al., citing Gobl and Ní Chasaide (1988), claim that preaspirated plosives may start as assimilation to the breathiness of the preceding vowel. Further anecdotal evidence in support of the link between breathy vowels and preaspiration comes from Ladefoged and Maddieson (1996). They report that the one speaker of the Arawakan language Guajiro that they recorded “sometimes used a breathy voiced offset to a vowel that was followed by a long stop, but he did not pre-aspirate consistently” (p.72). So it may be that the articulatory gestures that result in preaspiration are the same as those that result in breathy vowel offsets. If that is the case, then sex-based differences in preaspiration are related to sex-based differences in aspiration.

Helgason et al. (2003) demonstrate that the sex-based differences in preaspiration in Swedish reported in the previous paragraph are a change in progress. They present data from twelve speakers of Swedish (three younger males, three younger females, three older males, and three older females) that showed that younger people use less preaspiration, suggesting a change in progress from longer preaspiration durations to

---

4 We will see a similar tendency in English in the discussion of engendered phonetic-level variation and language change (§4.4).
shorter preaspiration durations. Furthermore, men preaspirate for shorter durations than women regardless of age, so males are leading the change. Helgason et al. point out that in other studies of language change in Swedish, young females are always in the lead. The case of women leading sound change is well observed, and enshrined as a principle of sound change (Labov 1990, 2001). This apparent exception to the principle, along with other similar examples, will be discussed in Chapter Four. The researchers assume “a partly biological and partly learned basis” (p.20) for this apparent exception to the principle, without further comment.

These studies show that regardless of language spoken, women produce more breathy voice quality than men, and this breathy voice quality is due to differences in the behaviour of the vocal folds, viz., female vocal folds tend to vibrate with incomplete closure, while male vocal folds tend to vibrate with complete closure.

2.5.2 Other studies of breathiness

Henton and Bladon (1985) measured the breathiness of vowels for 61 English speakers from two dialects in England, Modified Northern and Received Pronunciation. They calculated the noise component of the speech signal for each speaker by comparing the ratio of intensity of the first harmonic to the second harmonic. They found that women had much breathier vowels than men, and that Modified Northern speakers were more breathy than Received Pronunciation. They claim that women use a more breathy voice to create a feeling of intimacy, and suggest that there is a biological explanation (p.226):

A breathy voice is a ‘sexy’ voice (Crystal, 1975, p. 85; Daniloff et al. 1980, p. 175) and a sexy voice is associated with arousal. It may not be too speculative to assume a physiological basis for the association between breathiness and arousal. An accompaniment to the release of sex hormones by the hypothalamus is the release of other secretions—lubrication to the body as a whole, and the larynx does not escape this effect. If the larynx receives extra lubrication, then this may inhibit the ability of the vocal folds to adduct fully, resulting in an inefficient phonation and producing breathy voice.

Though speculative (they do not actually measure the amount of lubrication of the larynx during a state of arousal), this is another possible explanation for the sex differentiation pattern seen in the use of breathy voice quality, and it may well be that incomplete closure of the vocal folds is an attempt by women to imitate the lubrication effect.
Henton and Bladon do not discuss the dialect differences (but see Henton and Bladon 1988).

Ferrand (2000) examined the HNR in four age groups of children. The results did not show a clear pattern of sex differentiation, but did reveal that children’s voices are noisier than adults. She concludes that the results from the studies done on adults do not apply to children.

Ito (2003) looked at breathiness and politeness in Japanese. She recorded five Japanese males, and then did a perception test with listeners judging how polite the voices are. She found that the bandwidth of the first formant correlated significantly with the politeness judgments. She also points out that another marker of politeness (albeit “negative” politeness), high-pitched voice, is avoided by Japanese men.

Linville (1992) looked at glottal closure via videostroboscopy for two age groups of English-speaking women, one group about 21 years old, and the other about 80 years old. Each speaker produced sustained /i/ vowels using high, normal, and low pitch and a soft, normal, and loud voice. She found that the young and elderly groups do not differ significantly in the overall incidence of glottal gaps (incomplete closure of the glottis during the close phase of the phonation cycle). However, further examination by type of incomplete closure did show differences. The relevant types are:

1. complete: glottis is completely closed during each close phase of the cycle
2. posterior chink: vocal cords are approximated except in the cartilaginous portion of the glottis
3. anterior gap: closure is incomplete in the anterior segment of the glottis
4. spindle: closure is incomplete in the membranous glottis from the vocal processes to the anterior commissure

Young women produced significantly more incidences of posterior chink, and significantly fewer incidents of anterior gap or spindle. Anterior gap was the most common type of gap for the older women. Linville points out that for the younger speakers to have significantly more incidents of glottal chink than the elderly speakers is surprising given the anatomical data on age-related changes of the larynx; older women
have less motor control in the larynx, and therefore should produce more incidents of incomplete closure, regardless of type. She suggests that “less forceful contraction of the interarytenoid muscles is adopted by younger women as an economic measure or as a means to accomplish a vocal aim, such as a slightly breathy voice” (p.1215). This last point suggests that speaking with a breathy voice (and by extension sounding more feminine) is more important to the younger females than to the older females.

These studies show that breathiness voice quality is produced more often by women than by men, and that this sex difference in production is grounded in differences in the behaviour of the vocal folds. Although the link to anatomy/biology is speculative, the direction of the sex difference is always the same, with women producing more incidents of breathiness. On top of this trend is the tendency for some speakers to use breathiness to indicate speech style such as polite speech. The consistent trend across languages, the gradient nature of breathiness, and the control individuals have over the production of breathiness make breathy voice quality a source of engendered variation.

2.6 Creak

2.6.1 Creak as engendered phonetic-level variation

Creak (also called vocal fry) is the process of phonating with a very low frequency of vibration, approximately 20–70 Hz (Henton and Bladon 1988:9), well below that of modal voice. It is produced by bunching vocal fold thickness, relaxing vocal fold tension, and reducing subglottal pressure (*ibid.*). The auditory impression of creak is a quick succession of popping sounds.

From the perspective of engendered phonetic-level variation, creak is interesting in that men and women produce fundamental frequencies within the same range, suggesting that there is a lack of a gender effect. Henton and Bladon (1988) propose the following explanation for this lack of between-sex differences in F0 (p.8):

If the vocal folds remain thick and bunched (cf. the photographs in Ladefoged 1982:128 [2001:123]), it may be that female vocal folds may assume this position as readily as males’ because the length and volume dimensions (on which male and female vocal folds differ considerably) are largely irrelevant to creak production. The essential bunching, however, can be attained irrespective of gross structural differences.
If what Henton and Bladon say holds, that is, female vocal folds may assume the position required for creaky voice as easily as males’, then it is debatable that creak is engendered variation since one of the key characteristics of engendered variation is that men and women are biased towards a specific phonetic-level production for anatomical and physiological reasons. The comments made by Henton and Bladon suggest that in the case of creak, the anatomical bias is missing.

Yet, Henton and Bladon themselves pointed out that of the three experiments they review that use both men and women, in two cases the researchers had difficulties with the female subjects not being able to reliably produce creak phonation (p.13; the two experiments are Hollien and Michel 1968 and Monsen and Engebretson 1977). They then commented (p.13, emphasis theirs), “we would suspect that if female speakers have difficulty producing creak even when asked to do so, it is because creak is a phonational register not commonly employed by them.” The researchers showed experimentally that men produce creak more often in two dialects of British English, Received Pronunciation (RP) and Modified Northern (MN). The Modified Northern dialect is defined as speakers who grew up in or around Leeds, but have moved away from the area for substantial periods of time (for example, for work or university), and thus have modified their accent. They measured the percent of syllables that had audible creak in several sentence positions for men and women. They found that women use very little creak at all regardless of dialect, and that men who spoke the MN dialect use a lot more creak than men who speak RP. This shows that creak is both a marker of gender and of dialect.

This raises the question: Why do men more easily and more often produce creak than women? The answer to this can be found in the more recent work on the physiological and aerodynamic differences of the vocal folds of men and women during the production of modal phonation register and creaky phonation register.

One observed difference between the glottal waveforms of men and women is the open quotient. Figure 2.7 shows the various measurements taken from the glottal cycle. Open quotient is defined as the ratio of the opening time to the overall period of the glottal cycle \([t_1 + t_2]/T\). Holmberg et al. (1988) made several measurements on the inverse filtered airflow waveform of productions of syllable sequences in soft, normal, and loud voice for 25 male and 20 female speakers, including fundamental period \(T\),
open quotient, speed quotient (the ratio of opening time to closing time $t_1/t_2$), peak flow, minimum flow (or “dc offset”), and peak ac flow. These terms are borrowed from the field of electricity and are abbreviations of “alternating current” and “direct current.” They found several significant male-female differences in the waveform parameters. In comparison to women, the waveforms of men had a significant longer period, a higher peak flow, a higher ac flow, a larger speed quotient, and a smaller open quotient. This last finding implies that the close time ($t_3$ in Figure 2.7) of the cycle is longer in men than women. The relevance of this will become apparent shortly.

Another difference is the glottal waveform shape. Monsen and Engebretson (1977) used a reflectionless metal tube as a terminal point of the vocal tract in order to collect glottal volume-velocity waveforms from ten men and ten women. Each subject produced sustained schwa-like vowels using loud, normal, soft, falsetto, and creaky voice. They found that the waveform produced by the males during normal voice phonation was typically asymmetrical, and often had a hump during the opening phase of the wave (Figure 2.8). The waveforms produced by the women, on the other hand, tended to be symmetrical, and seldom had a hump during the opening phase (p.986). Although these sex differences occur during normal phonation, their relevance will become apparent shortly. So we see that the male glottal wave, besides being longer through the closed period than the waveform of females, is also typically asymmetrical and frequently shows a hump in the opening phase.

Monsen and Engebretson suggest that the different shapes in the glottal waveforms
are due to anatomical differences in the vocal folds (p.992):

It is possible that the asymmetrical, humped appearance of the male glottal wave may be due to a slightly-out-of-phase movement of the upper and lower parts of each vocal fold. If this is so, then the generally symmetrical appearance of the female glottal wave may be due to the fact that the shorter female vocal folds come into contact with each other more nearly as single masses.

Models of vocal fold behaviour approximate the vocal folds with two masses connected together by a coupler (see Stevens 2000:Figure 2.2). The two masses are free to move independently of each other, and their different sizes insure that their movement pattern are not in complete synchronization with each other. Monsen and Engebretson suggest that the more asymmetrical shape of male glottal wave form may be because the movement of the upper and lower sections of the male vocal fold is less synchronized than the movement of the upper and lower sections of the female vocal fold.

Besides modal register production, Monsen and Engebretson also looked at the production of creak. They found that the glottal wave during creak had a steeper rise and fall, followed by a long close period ($t_3$ in Figure 2.7), a trait shown by the male glottal wave form during modal production, and a large amount of period-to-period variation in the fundamental frequency. In their discussion of the creaky voice they do not point out any sex differences, but they did comment that some speakers produced a glottal waveform with two peaks per cycle (Figure 2.9).

Now consider the shape of the glottal wave form during creaky register. Blomgren et al. (1998) took acoustic, air flow, air pressure, and electroglottalgraphic and airflow...
measurements from ten men and ten women producing sustained vowels in both modal and creak registers. Although they found a significant sex difference for modal fundamental frequency, creak fundamental frequency did not show a significant sex difference. The air flow data showed that both males and females exhibited higher airflow rates in modal register than in creaky register, and that males had significantly greater airflow rates than females in modal register only. Air pressure did not show a significant effect for sex. So, although there are aerodynamic differences during modal register, these differences disappear when men and women produce creak. Similar to Monsen and Engebretson, the researchers found that many of the speakers produced creak glottal waveforms with multiple peaks per cycle. A breakdown by sex showed that for the male speakers’ creak productions, 83 percent of the waveforms contained multiple peaks. The women, on the other hand, produced multiple peaks only 38 percent of the time. Blomgren et al. do not offer an explanation for these observed sex differences. For that we need to turn to Chen et al. (2002).

Chen et al. (2002) recorded electroglottalgraphic waveforms from five men and five women producing sustained vowels in both modal and creak registers. From the
waveforms, they measured opening time, closing time, and speed quotient. They used a slightly different definition for speed quotient, in this case lumping $t_2$ and $t_3$ in Figure 2.7 together as the closing time. They found that both men and women had significantly greater speed quotient values in creaky register than modal register, and that women had significantly greater speed quotient values than men, regardless of register. As the speed quotient increases, the ratio of opening time to closing time increases. The sex difference in the speed quotient is due to the fact that while the male speakers show an absolute increase in both the opening and closing time durations during the production of creak, the women show only a slight increase in the closing time duration and a large increase in the opening time duration during creak as compared to modal voice (p.827). So it seems that the transition from modal to creaky register is a greater alternation of the female glottal wave, and by inference, the female vocal folds, than the male’s.

Chen et al. also found that men produced more multiple peak cycles (they call it a “dicrotic pattern”) compared to women: in this case 62 percent versus 20 percent, numbers comparable to Blomgren et al. These two findings — the sex differences in the speed quotient and the sex differences in the production of multiple peak glottal wave patterns — led the researchers to the following conclusion: “men exhibit greater vocal fold compliance during vocal fry phonation than women” (p.827). This claim is further supported by the discussion of the sex differences during the production of modal voice by Monsen and Engebretson (1977), and exemplified in Figure 2.8. The “hump” seen more frequently in the opening phase of the male’s glottal wave form during modal production is because of the asymmetries in the behaviour of the upper and lower vocal fold. It seems that this effect is amplified during creaky voice production, leading to more multiple peak cycles in the glottal waveform of males.

Together, these observations of sex-based differences in the production of modal and creaky phonations suggest that the male vocal folds more readily produce creak than female vocal folds for physiological reasons. This conclusion is reflected in the results of Henton and Bladon discussed at the beginning of this section, who found that both men and women produce creak during natural speech, but men do so much more often. We have now demonstrated the two of the characteristics of engendered phonetic-level variation, viz., that sex differences are founded on anatomical and physiological
differences in the vocal tract, and that these differences are reflected in the speech of men and women. In the next section, I will add a study that demonstrates the malleability of creak vocal quality, another essential characteristic of engendered variation.

2.6.2 Other studies of creak

In this section, I briefly introduce several studies that show that creak correlates with social factors other than sex. We saw previously that creak varied by dialect region (Henton and Bladon 1988). The following studies show! the variation of creak by social groups within the same dialect boundaries.

Esling (1978) looked at the voice quality correlations with social class in Edinburgh, Scotland. The author did impressionistic measurements of 34 voice quality characteristics, including creak, collected from 52 male speakers distributed across three status groups, high, middle, and low. He found that the high status speakers used a more creaky voice setting, while the low status speakers used a more harsh voice setting. Trudgill (1974), on the other hand, found that in Norwich, England, working-class speakers employ a creaky voice phonation setting, which was lacking in the speech of his middle-class informants.

Stuart-Smith (1999) gathered conversational recordings for thirty-two speakers, with equal numbers of men, women, older adults, young teenagers, from a middle-class neighbourhood and from a working-class neighbourhood. Each speaker was rated on several impressionistic voice quality scales, including creaky voice. She found that overall, Glaswegians made liberal use of the creaky voice setting, with men producing more creak than women. There was very little differentiation among the different male groups. Among the females, the working-class women and the middle-class teenagers produced more creak than the other female groups.

Redi and Shattuck-Hufnagel (2001) looked at different types of glottalization by sentence position by men and women. Although they include in their review of the literature studies on glottal replacement and glottal reinforcement of voiceless plosives, they do not include this type of variation in their definition of glottalization for their experiment. Instead, they identified four types of voice settings that they consider to be glottalization: periodicity, irregularity in duration of the glottal pulses from period to
period; creak, prolonged low fundamental frequency accompanied by damping of the pulse; diplophonia, repeated alternation of shape, duration, or amplitude of the glottal periods; and glottal squeak, a sudden shift to a relatively high, sustained fundamental frequency. They used read speech by professional newscasters and non-professionals. They found that the professional women produced the most glottalization, followed by the non-professional men, then the non-professional women, and the professional men last. We also see the same pattern when just the tokens that demonstrated the characteristics for creak are counted (Figure 2.10). The expected pattern of men using creak more often than women is seen in the non-professionals, but not in the professionals. The authors of the study do not comment on the reversal in sex differentiation pattern of glottalization between the professionals and the non-professionals. It seems that the women are performing hyper-masculine speech style, overshooting even the non-professional males.

From these studies, we see that creaky voice quality shows all of the characteristics of phonetic-level engendered variation. Due to differences in the make-up of the vocal folds of men and women, men are more likely to produce a creaky voice quality. Although all of the studies discussed here are on the English language, we can assume that men and women have the same vocal fold make-up regardless of dialect or language,
and that therefore this trend will hold across languages. At the same time, creak is gradient in nature, so that some social groups produce more creak than others. Furthermore, there are reversals in the trend of men producing more creak, such as the female professional newscasters seen in Figure 2.10. The consistent cross-linguistic trend and gradient nature of the sex differentiation pattern are two of the essential characteristics of phonetic-level engendered variation.

2.7 Summary of the characteristics of engendered variation

In this chapter, I have introduced several examples of phonetic-level engendered variation. The examples include both segmental variation (VOT, sibilant centre of gravity) and suprasegmental variation (average F0, breathiness, creak). For each case, we have seen evidence that the variation between the sexes seems to be grounded in the anatomical differences of the vocal tracts of men and women. These differences in the vocal tract are reflected in our speech in terms of tendencies for a person of a certain sex to speak a certain way. Specifically, men, in comparison to women, tend to produce shorter average VOTs, lower frequency sibilant Centres of Gravity, lower average F0, fewer instances of breathy voice, and more instances of creaky voice. These sex-based differences in our speech seem to be a product of differences in the anatomy of the vocal tracts of men and women, and as these anatomical differences are universal, the patterns shown by the variation should show a consistent trend from one speech community to the next.

At the same time, phonetic-level engendered variation is capable of acting as a sociolinguistic variable in that it marks group membership, be it ethnic group, class group, gender group, or speech style. As well, individuals, or even specific groups of speakers, may show exaggerated, reduced, or in some cases, reversed patterns. These points are summarized as follows:6

6 All of the hypotheses introduced in Chapters Two, Three and Four are gathered together for ease of reference in Appendix A.
There are two types of engendered phonetic-level variation: masculine and feminine. The basic, underlying trend for a masculine variant is that men tend to produce it more often than women for anatomical reasons. One example of a masculine variant is voicing. In the same way, the basic, underlying pattern of a feminine variant is that women tend to produce it more often than men for anatomical reasons. An example of a feminine variant is breathy voice. These labels also apply to variants of the same variable. For example, we can say that low sibilant centre of gravity is masculine engendered phonetic-level variation, since while sibilant centre of gravity varies from person to person and speaker group to speaker group, on average men tend to produce lower sibilant centres of gravity than women. In the same respect, higher centre of gravity is feminine engendered phonetic-level variation.

The distinction between masculine engendered variation, feminine engendered variation, and nongendered variation will be useful throughout the rest of the thesis. However, it is important to keep in mind that the distinction only refers to the basic trend,
and that the trend may be overridden for various other reasons such as the social gender of the individuals, prestige, local norms, etc.
3 Behavioural Sources of Engendered Variation

The linguistic system in a speaker’s mind / brain is deeply unconscious… in the same way that our processing of visual signals is deeply unconscious.

Jackendoff 2004:652

3.1 Phonetic distinctiveness

In the last chapter I gave several examples of phonetic-level engendered variation. For each of the examples, I argued that the sex differences were grounded in anatomical differences in the vocal tract of men and women. In this chapter, I extend the scope phonetic-level engendered variation to also include sex-based differences that are clearly not grounded in the anatomical differences, and yet otherwise still show the same characteristics as anatomically-grounded phonetic-level engendered variation. I label this type of variation as “behavioural.” An example of behavioural-determined phonetic-level engendered variation is that women articulate their speech more carefully, resulting in, for example, greater distinctions between vowels. This sex-based difference was observed early on by sociolinguists such as Labov (1972:304):

On Martha’s Vineyard, men are more ‘closed-mouthed’ than women, and use more contracted areas of phonological space; … women in New York City and Philadelphia use wider ranges of phonological space than men.

I gave other examples of researchers remarking that women articulate more carefully than men in the discussion of VOT in the previous chapter (e.g., Scharf and Masur 2002, Whiteside and Marshall 2000).

Phoneticians, similar to the sociolinguists, most often remark on articulation differences between men and women with regard to vowel space differences, and that is where our discussion will start. This is followed by vowel and consonant reduction in
§3.3, phonemic vowel length contrasts in §3.4, and allophonic variation in §3.5. The objective of these sections is to illustrate a diverse range of phonetic-level phenomena involving “phonetic distinctiveness.” I define phonetic distinctiveness as the acoustic or temporal distinction between the phonetic realizations of two related phonological constituents. The relationship between the phonological constituents can be either paradigmatic or syntagmatic. In Chapters One and Two we saw that women tend to have a greater VOT difference between the voiced and voiceless categories in English, amplifying the distinction between the phonological categories of voiced and voiceless. This is an example of the phonetic distinctiveness of two phonological constituents (the voiced and voiceless features) in a paradigmatic relationship. Phonetic distinctiveness also applies to differences between a segment and its surrounding environment (a syntagmatic relationship), and between two allophones of the same phoneme (a combination of paradigmatic and syntagmatic relationships). These will be illustrated shortly.

The rest of the chapter is dedicated to reviewing previous studies of behaviourally-determined engendered variation. For each study, I point out the phonological contrast and the corresponding phonetic distinction examined. If the generalization that women speak more clearly than men holds for all speech situations (dialect, language, speech style, etc.), then we expect to see that in every case the phonetic distinctiveness of the phonological contrast produced by women is greater than that produced by men. And as we shall see, in every case the female speakers as a group consistently articulate speech in a manner that amplifies the distinction. At the end of this chapter I present the hypothesis that these behavioural sex differences are engendered at the level of phonetics. By this, I specifically mean that clearer enunciation in the form of greater vowel distinctions, less vowel reduction, etc., has social significance.

3.2 Vowel space

Phoneticians initially thought that vowel formant differences between men and women could be explained by the differences in vocal tract size. They set out to find a mathematical formula that would normalize the vowels and thus eliminate any sex differences. However, in 1966 Ignatius Mattingly presented counterevidence to this normalization hypothesis at the 71st meeting of the Acoustical Society of America
(Mattingly 1966). He concluded: “the separation between male and female distributions for some vowel formants is much sharper than variation in individual vocal tract size can explain” (p.1219). Fant (1975) came to the same conclusion. He found that the difference between the formants of men and women from several languages showed so much variation by vowel and formant number that any normalization technique would need to be both formant- and vowel-specific.

Some progress was made by Bladon et al. (1984), who noted that much of the variation seen in Fant’s data is eliminated if you use a non-linear scale such as Bark instead of the linear Hertz scale that Fant used. The non-linear scale is motivated by the way we perceive pitch, the perceptual correlate of frequency. Bladon et al. found that when female formant values for Modified Northern British English were scaled down about one Bark, they matched on average the male values quite closely. However, when they expanded their study to include both their own data and others’ data for other dialects and languages, they found that the scaling factor of one Bark varied from language to language (Figure 3.1). Here we see that Received Pronunciation British English shows the greatest sex differentiation in vowel formants, while Utrecht Dutch shows the least. Bladon et al. comment that the variation in the degree of sex

![Figure 3.1 Scaling factors for male / female vowel normalization for seven languages and dialects. From Bladon et al. 1984, Figure 6](image-url)
differentiation “come[s] as no great surprise, since it is axiomatic that the social structure of communities varies, and this will be reflected in speech” (p.67).

There is, of course, no reason to suspect that the pronunciation of vowels by women is any more or less clear than the pronunciation of vowels by men from Figure 3.1. This came slightly later when Henton (1995a) demonstrated that the “variation” seen between the males’ and females’ formant values is systematic so that females’ vowel spaces are uniformly larger than males’ vowel spaces. Henton plotted normalized vowel data for six languages and dialects on F1 by F2 plots using a Bark scale. Figure 3.2 shows plot for the Received Pronunciation dialect of British English. The data was normalized for comparison across sexes by subtracting one Bark from the females’ values. For all six plots, the females’ high vowels were higher, the low vowels were lower, and the back vowels more back than those of the males’. This same result was also found for Korean, even after normalization (Yang 1992, 1996). If the differences in vowel formants were solely due to differences in vocal tract size, then we would expect that all of the vowels would be offset in the same direction so that, for example, all females’ vowels would be higher and more front than the corresponding males’ vowels. Although it is not obvious from Figure 3.2, a comparison of all of the plots shows the differences tend to be greater in the F1 dimension. F1 inversely correlates with tongue height. From this we can conclude that the differences in the vowel plots are not anatomical since if anything, men have larger mouths than women, and therefore should be able to produce greater height distinctions than women. Rather, as Henton phrases it, women are simply more “open-mouthed” than men when speaking, which in turn implies clearly pronunciation via greater vowel height contrasts.

Henton suggests that there is a connection between the larger vowel space of women and Labov’s principles of sex differences in speech. One of Labov’s principles is that, “In stable sociolinguistic stratification, men use a higher frequency of non-standard forms than women” (1990:205). Henton assumes that the women in her study were aware that they had been selected as speakers of a standard variety of the language, and therefore did their best to produce ideal representations of that standard variety (p.421). This in turn resulted in a greater contrast between the peripheral vowels. Women produce more dispersed vowels than men and women use more standard variants than men, so
Henton assumes that a dispersed vowel space is more standard. I began this chapter with the comment that women articulate their words more carefully than men. So far we have seen that women, if nothing else, articulate their vowels so that there is a greater acoustic contrast. But are there real perceptual consequences of this? In other words, is the speech of women easier to understand? Bradlow et al. (1996) give empirical evidence that yes, the speech of women is easier to understand, and furthermore, vowel dispersion correlates with talker intelligibility. They recorded 20 talkers, ten men and ten women, reading five sentences each in English, and then presented these to naïve listeners who transcribed them. Each sentence was marked either “correct” or “has errors.” They then ran correlation tests between transcription accuracy and the global characteristics of speaker sex, speaker average F0, and rate of speech. Each talker’s vowel dispersion was also estimated by measuring the distance in F1 by F2 space from each production of the peripheral vowels /æ/, /o/, and /i/ to the centre of that talker’s vowel space (the average of all tokens), and then taking the mean of all the distances. They found a significant correlation for sex and transcription errors, with fewer transcription mistakes made for the sentences read by women. Within each sex, average F0 does not correlate with transcription accuracy. However, a greater range in F0 within the speaker did correlate with more accurate transcriptions, so it seems that more varied intonation makes sentences more intelligible. Speaking rate did not show a
significant correlation. Bradlow et al. found a moderate correlation between vowel dispersion and intelligibility: speakers with more dispersed peripheral vowels were transcribed with fewer errors (Spearman rank order correlation $\rho = 0.431$, $p = 0.060$). However, within-category dispersion (how tightly productions of the same vowel phoneme clustered together) did not show a significant correlation with the transcription error rates. This suggests that the degree of coarticulation of the vowel with its surrounding environment is not a factor in intelligibility. Bradlow et al. conclude that a talker’s vowel space is a good measure of overall intelligibility (p.270).

At the same time, there continue to be several researchers who pursue the quest for an explanation of the formant differences between men and women that is based only on our perceptual and physiological speech systems, and is therefore completely disconnected from any sociolinguistic explanation. Two such studies will be examined here (but see also Fant et al. 1991).

Diehl et al. (1996) argue that the expanded vowel space produced by female speakers is a product of their higher fundamental frequency. Since a vowel’s harmonics are integer multiples of the F0, a higher F0 entails more spaced out harmonics. Diehl et al. hypothesized that if the harmonics are more spaced out then there is a greater chance that a formant peak will fall directly between harmonics. The greatly reduced energy in the space between the harmonics will then result in the formant being ambiguous or harder to perceive. A larger vowel space compensates for this by creating sharper distinctions between vowels that are adjacent phonetically by further spacing out their formants. To test this hypothesis, they created artificial monotonic high lax vowels with increasing F0, so some of the tokens will have formants that fall evenly between harmonics, but others will not. Listeners then heard and labelled each token. Diehl et al. found that as F0 increases, the labelling accuracy decreases. They also found that if the F0 varies during the presentation of the vowel (e.g., falling intonation) then accuracy increased. They explained that this was due to the eventual overlap between harmonics and formants as the distance between harmonics was altered. This “fleshing out” of the spectral envelope resulted in the vowels with a dynamic pitch being easier to perceive than vowels with a static pitch. Their results certainly support their hypothesis, but their conclusion, “the greater between-category dispersion of female vowels may be plausibly explained as a
means of offsetting the deleterious effects on vowel identifiability” (p.205), is not infallible. Simpson (2001, discussed in the following paragraph) points out that normal speech situations do not contain stretches of monotone vowels, and a women’s voice often has a noise component, which would help fill in the gaps between harmonics and therefore compensate for the poor harmonic sampling due to a higher F0 (p.2154).

Simpson (2001) examined the hypothesis that the vowel space sex differentiation is due to structural differences in the vocal tract, and that an examination of the movement of the tongue in time will lead to an explanation of the differences. Simpson begins his discussion with four predictions concerning articulator speed, target undershoot, and vowel space, based on the fact that women have less distance to travel (smaller mouths) to reach the same articulator position as men:

1. The vowel durations are the same for men and women, and women show less undershoot since the target is reached in less time. The sex difference is greater for the open vowels and short vowels than for close vowels and long vowels, since the distance travelled is greatest for open vowels, and the time to reach the target is less in short vowels.

2. The undershoot of women is the same as men, but vowel duration is shorter since women can accomplish the same target as men in less time.

3. Undershoot and vowel duration are the same for men and women, but articulator movement is slower for women than for men since their tongues have less distance to travel to reach the target.

4. Undershoot and vowel duration are the same for men and women, but the women aim at targets that are farther away (i.e., bigger vowel space) than those of the men. The difference increases the more open a vowel is.

To test these predictions, pellets were attached to the tongues of men and women in several locations along the sagittal plane. The pellets’ locations were measured at several points in time during the pronunciation of the diphthong in the word light. The results of the pellet measurements show that women produce significantly longer diphthongs although the duration of the lateral segment and overall word length did not show significant differences; men move their tongues farther; and men move their tongues
That men move their tongues farther and faster than women during the articulation of the diphthong is in line with Simpson’s predictions. However, that women produce diphthongs with longer durations is not. Simpson begins his discussion of this unexpected difference by examining the possibility that the cause is what he calls “sociophonetic,” i.e., women produce longer segments because they speak more clearly. He dismisses this possibility for two reasons. First, the other durational measurements such as the duration of the word-initial lateral were not significant. If women were speaking clearer than men, then they should also produce the /l/ with a longer duration. Secondly, women also produce longer diphthongs and monophthongs in German (Simpson 1998), suggesting that the difference is not socially conditioned (which is specific to the language community), but rather is consistent across languages. If the pattern was socially conditioned, then we would expect to see variation in durations from one speech community to the next, but the German results are consistent with the English results. So it seems that regardless of the language, female durations are always greater than male durations (p.2163). He goes on to say that we need to look for a biomechanical account of these differences, and gives two possibilities. The first possibility, perhaps women are slower because they are more accurate (less variable), he rejects because the data do not support it. The second possibility, perhaps women are slower because they have less stiff tongues (stiffer material responds faster to external forces than less stiff material), he says is tenuous at best.

There are several problems with Simpson’s line of argument. First, he fails to point out that his reason for the differences seen in the diphthong duration do not account for women producing longer monophthongs in German as well. Since monophthongs consist of only one target, women should be able to obtain this target without requiring extra time to do so. Secondly, there is no reason why the “sociophonetic explanation,” that women articulate more clearly than men, could not remain consistent across languages. If the tendency for women to speak more clearly than men is consistent across languages, then the tendency for women to produce longer vowels should also be consistent across languages (although see the next paragraph). Secondly, if women do have less stiff tongues, and therefore require longer to articulate diphthongs, they should also require
longer articulating the other segments such as /l/, and we would expect, if anything, they would produce a smaller vowel space than men. If, instead, we think about this problem in terms of phonetic distinctiveness, as I suggest, then an explanation presents itself: women produce longer diphthongs because doing so creates greater phonetic-level distinctions, whereas producing longer /l/ segments does not (/l/ does not contrast with anything).

Furthermore, research on Swedish vowel duration shows that women do not always produce longer vowels. Ericsdotter and Ericsson (2001) measured the duration of stressed and unstressed vowels in read speech for five women and five men. Each subject read a sentence, after which the experimenter pretended to mishear one word in the sentence and asked the speaker to repeat it in order to elicit a different stress pattern. Ericsdotter and Ericsson found that after normalization women produced an average vowel duration that was longer for the men for stressed vowels, but if the vowels were not stressed then men produced the longer average vowel duration. These results suggest that women enhance the contrast between stressed and unstressed vowels to a greater extent than men do (p.36). These results are not in line with those of Simpson. This may be because, as Ericsdotter and Ericsson point out, Simpson did not examine vowel stress as a factor in determining vowel length. Regardless, if the differences in vowel length were a consequence of differences in tongue stiffness, then we would expect the effect to be consistent across different stress conditions. Ericsdotter and Ericsson make the absurd suggestion that “the experimenters were women using the female stress pattern in the ‘mishearings’ which may have made the interviewed women more cooperative and comfortable” (p.37). When the women are more comfortable, they use a less formal speech style. I say absurd because if you pause to think through their line of argument, you realize that they are implying that the women were more comfortable during the unstressed vowels only, and more stressed during the stressed vowels, since a longer vowel equates to a more formal speech style.

Further evidence for women producing a more dispersed vowel space comes from recent research on gay-sounding speech. Pierrehumbert et al. (2004), in a study of the influence of sexual orientation on vowel production, found vowel space dispersion differences based on both sex and sexual orientation. The researchers recorded 26 self-
identified heterosexual men, 29 self-identified gay men, 16 self-identified heterosexual women, and 32 self-identified lesbian / bisexual women reading a set of phonetically-balanced sentences in English. For each speaker, one token and /i/, /e/, /u/, /æ/, and /ɑ/ was extracted and the formants measured. The F1 and F2 formants for the five vowels were averaged together to give an approximation of the each speaker’s vowel space centre. Vowel space dispersion was calculated by measuring the Euclidian distance in F1 by F2 space from the each vowel to the speaker’s vowel space centre, and then averaged across speakers in a group. The heterosexual women showed a larger average overall vowel space dispersion compared to the men, 2.25 Bark for the women versus 1.93 Bark for the men. The group of gay men fell in between these two groups with an average of 2.12 Bark. Interestingly, the group of lesbian / bisexual women showed the largest vowel space with an overall average vowel space dispersion of 2.42 Bark. So again we see that women have more dispersed vowels than men.

As well, a number of studies have demonstrated that children begin to produce phonetic variation in vowels that is appropriate for their sex as young as four years old, long before they begin to show the sexual dimorphism in the vocal tracts that accompanies puberty. For example, Sachs et al. (1973) recorded twenty-six children ranging in age from 4 to 14 years, reading a single sentence. A phonetic analysis of these voices showed that the average vowel formant was lower for boys than for girls. The authors conclude that children are learning culturally-determined patterns that are appropriate for each sex. These results, viz., girls produce higher frequency formants than boys of the same age and stature beginning at a very young age, and have been reproduced a number of times (see for example, Bennett 1981, Bennett and Weinberg 1979a, Bennett and Weinberg 1979b, Busby and Plant 1995, Ingrisano et al. 1980, Perry et al. 2001, etc.). The vocal tracts of the young children do not show significant sex differences (Fitch and Giedd 1999), suggesting that the sex differentiation in the vowels shown by young children and then as adults later on in life is social in origin.

Either way, regardless of the plausibility of any explanation based on our perceptual and physiological speech systems such as those provided by Diehl et al. and Simpson, we cannot ignore the cross-linguistic consistency of the claim that women produce a greater vowel space than men, as shown by studies such as Henton (1995a)
and Yang (1992, 1996). Even if, as suggested by Diehl et al. and Simpson, this variation is grounded in the differences in physiology of men and women, the actual degree of compensation for those differences varies from one speech community to the next, as shown in Figure 3.1. So again we see the familiar pattern of engendered variation — a consistent cross-linguistic trend in the pattern of sex differentiation — that shows gradient variation at the level of the speech community.

3.3 Vowel and consonantal reduction

We now move on to a phenomenon that is undisputedly not related to anatomical differences in the vocal tracts of men and women: vowel reduction. Vowel reduction is a feature of stress languages such as English and Swedish. In these languages, stressed vowels are more intense, more peripheral, and longer than unstressed syllables. As well, in English as the degree of stress diminishes the vowel quality tends to approach that of the schwa (for example, Fourakis 1991, Lieberman 1960, Tiffany 1959, etc.). None of the vowel qualities just mentioned is a product of differences in the vocal tracts of men and women; both sexes are just as capable of, and do produce shorter, less peripheral, less intense vowels in natural speech. But the observation that women produce larger vowel spaces has motivated at least one researcher to investigate the hypothesis that men produce schwa-like vowels more frequently than women.

Byrd (1994) carried out a series of studies that examined phonetic-level variation in the TIMIT speech database. The TIMIT database was developed jointly by Massachusetts Institute of Technology, Texas Instruments, and SRI International. It consists of 2342 sentences of read speech by 630 speakers of American English. All of the sentences have been segmented and labelled with enough detail to show several different types of subphonemic variation. For each speaker, dialect region, age, sex, and ethnicity were recorded. Byrd examined the frequency of a number of different types of reduction by sex and dialect region. Besides vowel reduction, she also examined speech rate, release of sentence-final stops, flapping, glottalization, /h/ voicing, syllabic consonants, and palatalization. Table 3.1 summarizes her results, with the line relevant to our discussion in bold. Although all of her findings are relevant to the broader topic of reduction, I only discuss her vowel reduction findings here. I will return to one other finding in the table at the end of this section. Byrd examined vowel reduction by
counting the frequencies of the occurrences of the three central vowels [ɔ], [i], and [ʌ]. Byrd counted a total of 17,858 central vowels, of which 27 percent were [ɔ], 55 percent were [i], and 18 percent were [ʌ]. She did not find a significant effect for sex or dialect region on the frequency of the total number of central vowels, but when each vowel was considered individually, significant effects were found. Specifically, men used the schwa more often than women, and women used the other two variants more often than men. She also found a significant effect for dialect region of the use of [ɔ] and [i], but not [ʌ]. The uniting factor between [i] and [ʌ] is that, while all three vowels are central, [i] and [ʌ] are more peripheral than [ɔ]. We can interpret Byrd’s findings as follows: while both men and women reduce vowels by using a more central variant, men reduce vowels to a greater extent than women. That the different dialect regions show variation is worth commenting on. If the finding reflects actual dialectal differences, then we see that, similar to sex differentiation in vowel space discussed in the last section (particularly Figure 3.1), vowel reduction also shows variation from one speech community to the next, one of the characteristics on engendered variation listed at the end of the last chapter. However, Byrd doubts the dialect results, as the dialect effects may be an artefact of the differences in the ratio of male participants to female participants for each dialect region.

Another example of vowel contrast and vowel reduction comes from Henton (1990). Henton looked at variation in the formants of /ə/ in three dialects of English: Modified Northern, Received Pronunciation, and West Coast American. Here we will
only consider the results for West Coast American. She recorded five females and five males. Tokens of /ʌ/ and /ə/ were selected from consonantal contexts in a reading passage. For each token, the first and second formants were measured at the vowel mid-point. Figure 3.3 shows ellipses marking the first and second deviations of the variation in the formant values for ten tokens of each vowel from each speaker. The extent of overlap of the ellipses for the two vowels gives us an idea of degree reduction of /ʌ/. In the left-hand side of the figure, we see a large amount of overlap, indicating that the men reduce /ʌ/ to schwa to a greater extent than the women, who are shown on the right. In fact, the men seem to be almost merging the two vowels. Henton points out that the distinction between /ʌ/ and /ə/ has already been lost before /ɜ/ in most dialects of English (e.g., “fur” versus “fir”), and then speculates that this merger may be spreading to other environments. If so, men are leading the change. This is what we expect given Byrd’s conclusions that men tend to produce lax vowels as schwa more often than women.

These two studies of vowel reduction suggest that men reduce vowels more frequently, and to a greater extent than women. Unfortunately, our evidence for any claims of the cross-linguistic nature of the sex differentiation pattern seen in vowel reduction suffers from a lack of linguistic diversity, with the data for both studies coming from English. However, the direction of the trend, with women showing less reduction, and therefore a greater contrast than men between lax vowels, fits with the overall pattern of women producing sharper phonetic-level distinctions than men.

We can add one more study to this. Avery and Liss (1996, discussed in §2.3.1) examined vowel reduction in the more-masculine- and less-masculine-sounding speech. The researchers found that the voices that were perceived as more-masculine-sounding after controlling for fundamental frequency showed, among other cues, more vowel reduction. Again, the direction of the result fit the other two studies, and demonstrates that vowel reduction indexes perceptions of social gender.

I will now review two studies of consonantal reduction. Very little work has been done on sex differences in consonantal reduction, but the two studies reviewed here corroborate each other and one of Byrd’s results listed in Table 3.1. Both studies look at the production of word-medial /t/ and /d/. The two studies are Zue and Laferriere (1979) and Bauer (Bauer 2005). Zue and Laferriere recorded three male and three female
speakers of North American English reading words containing word-medial alveolar stops in carrier phrases. An acoustic analysis of the stops revealed sex differences. In most environments, the durations of the stops produced by the women were significantly longer than the stops produced by the men. Furthermore, not only were the male stops shorter, but men also produced fewer of the canonical form (an alveolar stop with oral constriction), instead producing flaps, glottal stops, voicing / nasal assimilation, or deleting the stop entirely.

Bauer examined variation in the lenition of the flap allophone of the alveolar stop in various environments in North American English. He recorded two males and three females reading sentences. Even though every token was auditorily perceived as a flap, some showed greater lenition than others. He empirically determined the degree of lenition for each flap by contrasting the intensity of the flap with that of the surrounding vowels. Heavily lenited flaps behave more similar to an approximant than to a stop, and therefore show smaller differences in intensity with the surrounding vowels than fully articulated flaps. Analysis of the data revealed that flaps tend to be lenited within words (e.g., created) more than across word boundaries (e.g., create it) for the men, but not for the women. In fact, the women seldom produced a lenited flap. Bauer points out that the sex differences in lenition cannot be due to physiological differences between men and women because the measure of lenition is not an absolute measure of intensity (which

---

Figure 3.3 Ellipses showing two standard deviations for the variation of the first and second formants of West Coast American English /æ/ and /o/ for five men (left) and five women (right). From Henton 1990, Figure 4
may differ due to anatomical sex differences), but rather is taken relative to the surrounding vowels. As such, the measure should abstract away from any differences in intensity that are due to anatomical differences.

One of the results from Byrd reported in Table 3.1 is that men produce the flap in flapping environments significantly more often than women (although Bauer found that all of subjects produced a flap every time it was a possibility). Collectively, these three studies suggest that there is a reduction continuum of production for the flap, with the non-flapped alveolar stop at the completely unreduced end, and the lenited flap at the fully reduced end. The less reduced the phoneme is, the greater the phonetic distinction with the surrounding environment. In other words, all three of these studies show that women produce more phonetically distinct alveolar stops / flaps than men in flapping environments. Unfortunately, theses studies are on one phoneme in one language, and therefore it is difficult to generalize the results to other segments and languages.

### 3.4 Phonemic vowel length

Phonemic vowel length contrast is another phenomenon that is undisputedly unrelated to the anatomical differences found in the vocal tracts of men and women. Many languages of the world make a phonological distinction between long and short vowels. For example, [koko] is the Japanese word for *here*, while [koːkoː] is the word for high school. The length contrast is realized phonetically in the temporal domain, with long vowels being perceptibly longer than short vowels. Based on the discussion of Simpson 2001 (§3.2), we might predict that women, with their “less stiff” tongues, will produce longer short vowels (recall that less stiffness results in longer time to reach the target), but produce long vowels of the same length as those produced by men. This is because in the case of the long vowel, the women have sufficient time to reach the articulatory target, and therefore do not need to unnecessarily extend the length of the vowel.

Johnson and Martin (2001) found that women actually produced longer long vowels than men in Creek, a Muskogean language spoken in Oklahoma and Florida. Creek has a three-vowel system, all of which are phonemically contrastive for length. Johnson and Martin measured the duration of short and long vowels produced in isolated words for four male and four female native speakers of Creek. They found that on
average, the ratio of long vowels to short vowels was about 1.8. Statistical analysis showed that vowel quality, vowel length (long versus short), speaker sex, and word position (word-initial versus word-final) all had significant effects on vowel length. Long vowels were longer than their corresponding short vowels, word-final vowels were longer than the vowels in word-initial position, and women produced longer mean durations for both distinctive vowel length categories. Furthermore, vowel length and speaker sex interacted, as shown in Figure 3.4, so that the between-sex difference is greatest for the long vowels. Johnson and Martin claim that “this interaction can be taken as a suggestion that the gender difference was not one of different speaking rates over all [which was not tested directly], but rather a difference in the phonetic realization of the length contrast” (p.87).

Johnson and Martin’s conclusion — women produce a greater temporal distinction between phonemically short and long vowels than men — is in the direction we predict, and adds further weight to the claim that women articulate speech in a manner that creates greater phonetic-level distinctions.

### 3.5 Allophonic variation

Another type of linguistic contrast is that between different allophones of the same phoneme. In this section, I give three examples of allophonic variation, one from Japanese, one from Arabic, and one from Canadian English. In the first two cases, as we shall see, the females speakers make a sharper contrast between allophones than the male speakers. This last case is an example of men producing more co-articulation than women.

#### 3.5.1 High vowel devoicing in Japanese

In the last chapter during the discussion of VOT as phonetic-level engendered variation (§2.2.2), I introduced Whiteside and Irving (1997) and their suggestion that the greater VOT voiced / voiceless difference (VOTVVD) produced by their female subjects supports the claim of “females using more carefully articulated speech” (p.207). We now turn to a study of Japanese high vowel devoicing that also observed women producing greater voicing distinctions than men.
Japanese has five vowels, /i u e o a/. In standard Japanese, the high vowels /i u/ typically fully devoice between two voiceless obstruents. Mid and low vowels also show devoicing in the environment, but not to the extent of high vowels. Sugiyama (2004) examines the VOT of both high and non-high vowels in a devoicing environment and a non-devoicing environment. She elicited near minimal pairs from 20 native speakers of Tokyo Japanese via word lists. The devoicing environment consisted of a word-initial CV syllable beginning with a /k/ and followed by /t/ in the following syllable. The non-devoicing environment consisted of a word-initial CV syllable beginning with a /g/ and followed by a post-alveolar tap in the following syllable. The speakers consisted of equal numbers of both sexes, as well as equal numbers from two age groups (21–26 year olds, and 53–65 year olds). As expected, Sugiyama found that the VOT values of a syllable containing a high vowel in the devoicing environment were significantly longer than the VOT of a syllable containing a high vowel in the devoicing environment. Furthermore, she also found that a comparison of the VOT Voiced Voiceless Difference (VOTVVD) revealed that women from both age groups produced a much sharper distinction between the high vowel and non-high vowel categories in the devoicing environment, as shown in Figure 3.5. All four groups of speakers produce a greater VOTVVD for the high vowels.
than the non-high vowels. A comparison of VOTVVD for high versus non-high vowels for each speaker group in Figure 3.5 shows that the difference between the vowel categories is greater for women than for men. This is another example of women articulating their speech in a manner that results in greater phonetic-level distinctions than the speech of men.

3.5.2 Vowel pharyngealization in Arabic emphatics

Many dialects of Arabic have a phonemic contrast between a set of plain coronal plosives and fricatives [t d s z] and an “emphatic” set, with secondary articulation in the form of pharyngealization [t̚ d̚ s̚ z̚]. This secondary articulation entails a narrowing of the pharynx during the articulation of the emphatic consonant, and is maintained during the articulation of the following vowel. Kahn (1975) examined the first and second formants of the low vowel /a/ following emphatic consonants and plain consonants. She recorded two sets of minimal pairs that differed only in the plain / emphatic distinction (one pair for [s]/[s̚] and one pair for [d]/[d̚]) read by five male and five female Arabic university students, and six male and five female American university students learning Arabic as a second language. Kahn found that in general, the tokens following an emphatic consonant were slightly lower and farther back than the tokens following a plain

![Figure 3.5](image-url)
consonant. Furthermore, the distinction between the emphatic and plain vowels was significantly greater for the Arabic females than for the Arabic males, as shown in Table 3.2. In order to emphasize this distinction, Kahn calculated what she calls an “emphasis factor,” which is an indicator of the extent of the contrast between emphatic and plain tokens, and is included in the table. The American students, who all learned Arabic from male teachers, do not show sex differentiation in the contrast between emphatic and plain tokens. Khan explains the behaviour of the Arabic women as follows: “Their speech appears more stylized as well as more conservative than Arab men (in classical Arabic these emphatic distinctions are made more clearly than in some of the dialects)” (p.46–7). However, this claim that Arabic women use speech forms that are more like classical Arabic has been refuted by a number of more recent studies that show the opposite: in Middle Eastern societies, men use the classical Arabic variant more often than women, whereas women use the prestigious colloquial variant more often than men (see Labov 2001:270 for a summary). This would seem to leave us without an explanation for the linguistic behaviour of the Arabic women in Kahn’s study. Rather than being a matter of conservativeness or innovation, I claim that their behaviour is a matter of phonetic-level distinctiveness between the two allophones of /a/. That Arabic women produce a sharper contrast between the allophones than men fits with the claim that in general women produce sharper phonetic-level distinctions than men.

3.5.3 Canadian Raising

The flip side of women producing sharper phonetic-level distinctions is that men do not produce as sharp phonetic-level distinctions. One possible source of this reduction in phonetic-level distinctiveness might be that men produce their segments with a greater degree of assimilation to the surrounding environment. In other words, it may be that men co-articulate more than women, particularly when co-articulation is not required by the phonology.

One example of men showing a greater degree of co-articulation than women comes from Chambers and Hardwick’s (1986) study of Canadian Raising in Vancouver and Toronto. Canadian Raising refers to the centralization of the nucleus of the diphthongs /ai/ and /aw/ to [AI] and [AW] before voiceless consonants in dialects of English such as Canadian English (Chambers 1973). Chambers and Hardwick carried out
sociolinguistic interviews with 30 Canadians, 18 from Toronto, and 12 from Vancouver. All of the subjects were middle class, and belong to one of three age groups: 12 year-olds, 22 year-olds, and 46+ year-olds. As well, each group consisted of an equal number of males and females. The nucleus of each token of /æw/ was impressionistically categorized as one of three degrees of frontness and an index score assigned as follows: 0) not fronted [aw], [aw]; 1) somewhat fronted [aw], [aw]; and 2) fronted [æw], [æw]. Individual speaker index scores were calculated by dividing the sum of the token scores by the number of tokens and multiplying by 100 in order to produce an integer number.

Chambers and Hardwick found that in both cities, the younger the speaker the more they used a fronted nucleus. When the results are further broken down by sex, they show that women are leading the fronting of the nucleus in all age groups, and in both cities. Figure 3.6a shows their results for fronting in Toronto and Vancouver with the two cities averaged together. (The results for Toronto and Vancouver are similar, so averaging the two together does not skew the figures.)

At the same time, they also found that some of the Vancouver speakers produced a variant that was not found in the Toronto sample: a diphthong with a rounded nucleus [ow]. A count of the percentage of tokens before voiceless consonants with a rounded nucleus in the Vancouver data revealed that the male speakers produced the rounded

<table>
<thead>
<tr>
<th></th>
<th>[s]</th>
<th>[s']</th>
<th>[d]</th>
<th>[d']</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1</td>
<td>F2</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>Arabic males</td>
<td>780</td>
<td>1560</td>
<td>860</td>
<td>1340</td>
</tr>
<tr>
<td>Arabic females</td>
<td>800</td>
<td>1920</td>
<td>960</td>
<td>1540</td>
</tr>
<tr>
<td>American males</td>
<td>783</td>
<td>1467</td>
<td>892</td>
<td>1216</td>
</tr>
<tr>
<td>American females</td>
<td>840</td>
<td>1560</td>
<td>840</td>
<td>1410</td>
</tr>
</tbody>
</table>

Table 3.2 Average first and second formant values in Hz and emphasis factor E in italics for /a/ by preceding consonant. E = (F2-F2e) + (F1e-F1), where e = emphatic consonant. From Kahn 1975, Table II
nucleus variant notably more often than the female speakers, as shown in Figure 3.6b. The researchers explain the different sex patterns as follows. The fronting process seems to be one of standardization with American English, with the variants used by the youngest Canadians indistinguishable from the variants used by their neighbours south of the border (p.41). The Rounding process, on the other hand, is a local change limited to Vancouver. Chambers and Hardwick claim that there is no phonetic counterpart to it in either Canadian or American English, and therefore the change can not be one of standardization. Rather, it is “the result of straightforward assimilation of the unrounded onset vowel to the rounded off-glide” (p.42). It is also an example of men producing more co-articulation than women, which agrees with the general claim of this chapter that women produce sharper phonetic-level distinctions than men.

3.6 Putting it all together

In this chapter, I have argued that there is a second source of engendered phonetic-level variation: the behavioural pattern of women enunciating their speech more clearly than men. At the end of the last chapter I listed several characteristics of phonetic-level
engendered variation. The first was that when enough speakers are looked at, sex-based patterns emerged. In this chapter, we saw three such patterns at the phonemic level related to vowels: women have larger vowel spaces than men in many languages of the world, men reduce English unstressed vowels to schwa more frequently than women, and women produce a greater temporal distinction between Swedish stressed and unstressed vowels and Creek short and long vowels than men. At the allophonic level, we saw that women produce a sharper distinction between voiced and devoiced high vowels in Japanese, and pharyngealized and non-pharyngealized low vowels in Arabic. Lastly, we saw that women produce a sharper distinction between the nucleus and the glide in diphthongs in Canadian English. We also saw limited data on consonants that showed that women reduce the flap in North American English less than men. Collectively, these studies all point to the same general conclusion — women produce sharper phonetic-level distinctions than men.

The second characteristic of engendered variation is that these patterns are consistent from one speech community to the next. Due to the sparseness of studies for each individual phenomenon, it is difficult to claim that any of them hold across languages. The strongest evidence for a claim of cross-linguistic consistency comes from the vowel space data, where the same pattern of women producing larger vowel spaces was seen in both several Western European languages and Korean. If we consider the more general claim that women produce sharper phonetic-level distinctions than men, then the diversity of the linguistic phenomena covered in this chapter strongly suggests that this generalization holds cross-linguistically.

The third characteristic of engendered variation is that the sex-based patterns seem to be grounded in the anatomical differences in the vocal tract. This point needs to be modified so that we also include the behavioural differences introduced in this chapter. These differences are clearly not motivated by anatomical differences in the vocal tract, and therefore it may seem that there is no connection between Chapters Two and Three. The unifying force is the cross-linguistic consistency of the linguistic patterns introduced in these two chapters. The cross-linguistic consistency, viz., when enough speakers are looked at sex-based patterns emerge, is the key characteristic of engendered variation and underlies all of the assumptions of the theory of engendered phonetic-level variation.
Cross-linguistic consistency is a characteristic of the patterns introduced in Chapter Two. I also claim that it is a characteristic of the patterns introduced in this chapter. Although this claim is extremely strong, the evidence is supportive — I have yet to find a published counterexample to the claim that women articulate their speech in a manner that shows greater phonetic-level distinctions than men, regardless of dialect or language.

The list of characteristics of phonetic-level engendered variation provided at the end of Chapter 2 was based on anatomically-grounded variation only. Following is the same list, but modified to include the possibility of the behavioural sources of sex-based variation introduced in this chapter:

- when enough speakers are looked at, sex-based linguistic patterns emerge
- these patterns of engendered variation tend to be consistent across different speech communities
- many of these patterns are grounded in anatomical sex differences in the vocal tract
- other patterns are clearly not grounded in anatomical sex differences in the vocal tract, and therefore are behavioural
- the extent of the sex differentiation seen in the engendered variation varies from one speech group / community to the next, but the direction the sexes pattern tends to remain the same
- however, there are individuals or groups who for social reasons pattern more like the opposite sex

Together, these first two characteristics make a prediction with regard to the behavioural sex differentiation: women articulate their speech so the phonetic correlates of two related phonological constituents will be farther apart in acoustic and / or temporal space than the phonetic correlates produced by men. This is our first hypothesis, and it is also the first of a series of hypotheses concerning “phonetic distinctiveness.” This hypothesis is also one of the four hypotheses that forms the impetus behind Chapter Five and Chapter Six, the other three being the second Gender Hypothesis (page 86) and the two hypotheses on phonetic distinctiveness and language variation and change (pages
Gender Hypothesis 1: The phonetic distinctiveness of the speech of women is greater than that of the speech of men.

Phonetic distinctiveness is defined as the acoustic and/or temporal distinction between the phonetic realizations of two related phonological constituents. The relationship between the phonological constituents can be either paradigmatic, syntagmatic, or both.

Given the ubiquitous nature of phonological contrasts in the languages of the world, Gender Hypothesis 1 is easily tested. For example, following is a list of examples of phonological phenomena and their corresponding phonetic correlates. In each case, we predict that the difference between the phonetic correlates will be greater for women.

- tone: average F0 on high tone syllables versus average F0 on low tone syllables
- mora: length of a heavy syllable versus length of a short syllable
- geminates versus non-geminates: length of occlusion for a single plosive versus the length of the occlusion for a geminate plosive
- plosive voicing: VOT of voiced versus VOT of voiceless plosive (see §2.2)
- palatalization of plosives: height of the second formant for a palatalized plosive versus height of the second formant for a non-palatalized plosive at the onset of the following vowel

etc.

This list is by no means meant to be exhaustive. But it illustrates that our first hypothesis makes a number of testable predictions. Furthermore, it unites a number of disparate phenomena. We have already seen that many researchers such as Simpson have looked for the source of the sex differences in the production of some of these phonological contrasts in the anatomical differences of the vocal tracts of men and women. By shifting the focus from the anatomical differences to behavioural differences, we seek a common
source of the sex differentiation pattern exhibited by these diversified phenomena.

At this point, we need to pause to consider: If a language uses phonetic information to convey non-linguistic information will women still show a greater range of use? A good example of this is pitch in English, or intonation. English intonation is used to convey such paralinguistic information as surprise. But unlike pitch-accent and tone languages, it is not usually used to convey phonological contrast. Henton (1995b) recorded five men and five women reading a text passage. Several measures of pitch dynamism failed to find significant differences between the two sexes. Assuming that women who use a tone or pitch-accent language show a greater distinction between the phonological categories (Gender Hypothesis 1), then the answer to the question is women only show a greater range if they are conveying a phonological contrast. This needs to be tested empirically, but that is beyond the scope of this dissertation.

The last two characteristics listed in the last chapter, viz., the extent of the sex differentiation seen in the pattern varies from one speech group / community to the next, and the fact that there are individuals who for social reasons pattern more like the opposite sex, were not adequately illustrated in the preceding discussion of behavioural variation. Although we saw that the extent of the difference for the vowel space differs from one speech community to the next (Figure 3.1), we do not have corresponding evidence for reduction of unstressed vowels to schwa, nor for vowel length contrasts. Furthermore, we did not see an example of cross-gender transition by an individual, that is, a man who patterns with the women or a woman who patterns with the men.

We did not see an example in the above discussion because all of the researchers treat the sex variable as dichotomous, and do not look at individual variation. In the last chapter, we were fortunate that several of the researchers presented their results broken down by individual speaker, allowing us to pick out speakers or groups of speakers who patterned with the opposite sex. I claimed that these individuals or groups pattern more like the opposite for sex social reasons, that is, anatomical engendered variation indexes social gender. The same should hold for behavioural engendered variation. This is our next hypothesis concerning phonetic distinctiveness: the distance between the phonetic correlates of a phonological contrast indexes social gender.
This hypothesis states that the sex-based behavioural differences introduced in Chapter Three will correlate with social gender. This implies that there is a degree of agency to these differences, so that, for example, the Latina gang girls discussed in §2.4 reduce the clarity of their speech by reducing the extent of the differences in the phonetic correlates of phonological contrasts in order to sound more masculine when talking about topics such as gang membership and fighting. As far as I know, Gender Hypothesis 2 has never been tested for behaviourally-determined phonetic-level engendered variation. This is one of the main objectives of this thesis. In Chapter Five, I empirically test that behaviourally-determined phonetic distinctiveness correlates with social gender with a study of the duration of contextually long and short vowels in the speech of eight male radio DJs.

*Gender Hypothesis 2: Phonetic distinctiveness indexes social gender.*
Engendered Variation and Sociolinguistic Theory

There are thousands of different voices in the world about us; and there are many potential voices in each of us.

Murphy 1964:4

4.1 Introduction

In the last two chapters I introduced two sources of engendered phonetic-level variation, anatomical sex differences and behavioural sex differences. The variation in the speech signal that results from these sex differences naturally correlates well with biological sex. We also saw a limited number of examples of this variation correlating with social gender. For example, feminine-sounding men tend to produce higher frequency sibilant centres of gravity. In this chapter, I discuss the way this variation interacts with three other major sociolinguistic independent variables: social class, speech style, and age (i.e., language change). Researchers such as Labov have made a number of claims concerning the way gender interacts with each of these independent variables. Two examples of such claims are that working-class speech patterns are associated with masculinity (Labov 1966:494), and that in linguistic change from below the level of social awareness, women use more innovative forms than men do (Labov 1990:215, 2001a:292).

However, as I mentioned in Chapter One, these claims about the sociolinguistic behaviour of men and women are made without taking into consideration the sex patterns due to anatomical and behaviour differences. In some respects, claims made by Labov and others conflict with the predictions made based on the sex patterns introduced in Chapters Two and Three. For example, if a change from below is the introduction of a phonetic-level masculine variant into the speech community (for example, voicing), will
women still lead the change, as predicted by Labov, or will men lead the change, as predicted by fact that men use masculine variants more often than women (Chapter Two)? The objective of this chapter is to review several of these claims in light of phonetic-level engendered variation and to phrase corresponding hypotheses such as the answer to the previous question. The hypotheses introduced in this chapter, along with the hypotheses from Chapters Two and Three, form the set of hypotheses introduced in this dissertation, and are summarized in Appendix A. Again, I emphasize that this list is not intended to be complete. There are other sources of linguistic variation that interact with gender but are not covered in this dissertation, such as language contact (see for example Winter and Pauwels 2000 for an overview of the literature on the role of gender in language contact and shift in Australia). Rather that producing an exhaustive theory of engendered phonetic-level variation, my objective for this dissertation is to produce enough evidence to justify my claim that the sources of variation discussed in Chapters Two and Three need to be recognized and taken into consideration when doing phonetic-level sociolinguistic research.

4.2 Engendered variation and social class

4.2.1 Working-class values and 'masculinity'

From the inception of the modern sociolinguistic research paradigm, researchers have claimed a link between working-class values and masculinity. Labov (1966), in his discussion of why the New York City accent is so unappealing to women but not to men, wrote: “One of the main factors which contribute support to the working-class speech pattern of the city is its association with the cultural norms of masculinity” (p.494). Labov’s claim is based on the observation that men tend to use a higher proportion of the local, non-standard variants than women of the same social class (Labov 1990), and that women felt more insecure about the way they spoke if they had a New York accent (Labov 1966:495).

Edwards (1979) demonstrated this association between masculinity and working class in an ingenious experiment involving the perception of the sex of children. He had listeners judge the sex of 20 prepubescent working- and 20 prepubescent middle-class ten-year-old children (10 boys and 10 girls in each group) by listening to their tape-
recorded voices. Previous work on the perception of children’s voices has shown that listeners are able to guess the sex of a child with above chance accuracy, but errors are still made. Each child read a short text passage. Listeners then guessed the sex of the child by listening to the recordings. As well, another group of listeners rated each voice on the following four traits using seven-point perceptual scales: rugged-delicate, low pitch-high pitch, masculine-feminine, and rough-smooth. Overall, the listeners were able to correctly guess the sex of a child 83.6 percent of the time, a result that is comparable to that obtained in other studies. Edwards’s results are noteworthy because of the pattern of the errors when listeners guessed the sex of a child incorrectly. Table 4.1 shows a breakdown of the number of errors by sex and class. Among the working-class children, the sex of the boys was identified more accurately than the sex of girls. When we look at the middle-class children, we see that this pattern is reversed; the sex of the middle-class girls was identified more accurately than that of the middle-class boys. While the most important determiner of how a child’s sex is perceived is the actual sex of the child, social class also plays a role. The results of this study indicate that children from a working-class background tend to be perceived as boys, whereas children from a middle-class background tend to be perceived as girls. The results of the perceptual scale evaluations give further support to this conclusion: the working-class children were perceived as having lower-pitch voices, rougher voices, and more masculine voices than the middle-class children.

Another source of evidence for the link between working-class speech and masculinity comes from Trudgill’s (1972) self-reporting data on the speech of Norwich, England. Trudgill gave half of his 60 informants a Self-Evaluation Test. The test consisted of the investigator reading aloud different pronunciation variants of the same word, and the informant indicating which variant they themselves tended to use the most. For each of these words, one of the pronunciations was considered prestigious and used by speakers of Standard English. When the reported usages were compared to the actual usages, Trudgill found a striking result: Norwich men tended to under-report their use of the standard variant, while Norwich women tended to over-report their use of the standard variant (p.187). Trudgill explains these results in terms of what he calls “covert prestige.” Covert prestige is the hidden value attached to the speech of the working class by males in general. When asked, the Norwich men do not admit it, so they are covertly
attaching value to the speech of the working class. Trudgill conjectures (p.183):

WC speech, like other aspects of WC culture, appears... to have connotations of masculinity... probably because it is associated with the roughness and toughness supposedly characteristic of WC life, which are, to a certain extent, considered to be desirable masculine attributes.

This belief that working-class men are ideal representations of masculinity along with the desire to sound masculine motivates male speakers, at least, to over-report their use of the non-standard pronunciation variants. In the case of the women it is the reverse. Women’s desire to sound less masculine and to present an upwardly mobile image motivates them to under-report their use of the non-standard variants.

This connection between working-class speech and masculinity predicts that from the perspective of engendered phonetic-level variation working-class men and women will use masculine variants (as defined in §2.7) more frequently than middle-class men and women. But why would middle-class males value “sounding masculine” less than their working-class counterparts? The answer to this question is that there are multiple types of masculinity, and that working-class men value indicators of “physical masculinity,” while middle-class men value “technical masculinity” (Connell 2005:55–6):

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Number of Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>boys</td>
<td>140</td>
<td>4 (2.9%)</td>
</tr>
<tr>
<td>girls</td>
<td>140</td>
<td>30 (21.4%)</td>
</tr>
<tr>
<td>Middle class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>boys</td>
<td>140</td>
<td>43 (30.7%)</td>
</tr>
<tr>
<td>girls</td>
<td>140</td>
<td>15 (10.7%)</td>
</tr>
</tbody>
</table>

Table 4.1 Number of errors made in the identification of children’s sex by social class. From Edwards 1979, Table 1

Heavy manual work calls for strength, endurance, a degree of insensitivity and toughness, and group solidarity... working men’s bodily capacities are their economic asset, are what they put on the labour market... Middle-class men, conversely, are increasingly defined as the bearers of skill... The new information technology requires much sedentary keyboard work, which was initially classified as women’s work (key-punch operators). The marketing of personal computers, however, has redefined some of this work as an arena of competition and power — masculine, technical, but not working class. These revised meanings are promoted in the text and graphics of computer magazines, in manufacturers’ advertising that emphasizes ‘power’
(Apple Computer named its laptop the ‘PowerBook’), and the booming industry of violent video games.

Working-class men value physical toughness, while middle-class men value technical aptitude, particularly with electronics. This implies that working-class men value speech and voice that makes them sound physically tough more than middle-class men.

How do male voices make men sound “tough”? According to Ohala (1984, 1994) sexual dimorphism in the vocal anatomy of humans is designed to amplify the physical superiority of men over women. After puberty, the male larynx is approximately 50 percent larger than the female larynx, resulting in longer male vocal cords, and thus a lower F0. Furthermore, the male larynx lowers in the throat during puberty, resulting in a male vocal tract that is approximately 20 percent longer than a female’s. Ohala (1984:14) cites evidence, which he dubs the “frequency code,” from animal communications that shows that “high F0 signifies (broadly) smallness, nonthreatening attitude, desirous of the goodwill of the receiver, etc., and low F0 conveys largeness, threat, self-confidence and self-sufficiency.” For example, a dog will produce a high-pitched whine to express submission, but a low-pitched growl to threaten. The longer vocal folds in combination with the longer vocal tract give the adult human male lower resonances than the adult human female. This in turn gives an inflated perception of physical largeness (much like a male peacock’s tail feathers) because listeners associate lower frequency sounds with larger body size. But our voices are only capable of expressing “physical masculinity,” that is masculinity that is valued by the working class.

These considerations leads us to our first hypothesis concerning social class and engendered variation: working-class speakers use more masculine variants than middle-class speakers, while middle-class speakers use more feminine variants than working-class speakers. because working-class speech is associated with physical toughness.

**Social Class Hypothesis 1:** Working-class speakers use more masculine variants than middle-class speakers, while middle-class speakers use more feminine variants than working-class speakers.
With this better understanding of the relationship between social class and engendered variation, let us now return to Labov’s claim about the connection between New York speech, working class, and masculinity quoted above. This conclusion was drawn based on nongendered variation such as r-dropping. The connection, however, between working-class speech and masculinity is indirect: both working-class speakers and males use the non-standard variant more often, and therefore there is an association between male speech and working-class speech. However, there is nothing inherently masculine about r-dropping. Unlike the variation introduced in Chapter Two, there is no anatomical reason why men should prefer to drop ‘r’. Phonetic-level masculine variations, on the other hand, are inherently masculine. Now consider this from the perspective of the female. Working-class jobs are much more engendered than middle-class jobs, with construction workers, mechanics, etc. tending to be men, and waitresses, hairdressers, etc. tending to be women. But physical toughness is certainly not an asset to your stereotypical hairdresser; if anything, it is a disadvantage. Do working-class women value phonetic-level engendered variation to the extent that working-class men do?

This is a difficult question to answer due to the lack of relevant phonetic-level studies. We can infer the answer from the results from one study, Stuart-Smith (1999). She recorded conversational speech and word list readings for 32 speakers, with equal numbers of males, females, older adults (40–60 years), younger teenagers (13–14 years), from a working-class neighbourhood and from a middle-class neighbourhood. An impressionistic analysis of voice quality by expert judges revealed that males made more use of a creaky voice quality, while females made more use of a whispery voice quality (p.215). Furthermore, in conversational speech, the working-class speakers distinguish themselves from middle-class speakers by using, among other voice settings, a whispery voice (but not creak). Stuart-Smith comments that the accent of the middle-class speakers can be defined almost entirely by the absence of the voice features such as breathiness used by the working class (p.219).

Slightly related to Stuart-Smith’s find of a lack of differentiation among middle-class speakers is the claim by Henton and Bladon (1995, 1988; reviewed in Chapter Two) that the RP accent, which is associated with middle-class speech, represents a voice of “perceived androgyny” (used in the sense of undifferentiated sexuality, cf. Biemans
ENGENDERED VARIATION AND SOCIOLINGUISTIC THEORY

2000:50–6 for a different usage), a claim previously made by Elyan et al. (1978). However, their study compares a regional accent (Modified Northern) with the prestige variety, and so tells us very little about working-class speech.

These studies suggest that working-class men and women show greater between-sex differences than middle-class men and women. Why this might be is not clear. It may be that the working-class men value “physical masculinity,” while working-class women value “physical femininity.” This conjecture forms the basis for the next engendered variation hypothesis.

This hypothesis claims that while both working-class and middle-class speakers make use of both feminine and masculine variants, the sex differentiation will be greater for the working class. This is illustrated in Figure 4.1 by two hypothetical examples, one of a masculine variant, and the other of a feminine variant, by social class and sex. The working-class male speakers use the masculine variant more often than middle-class male speakers, while working-class female speakers use the feminine variant more often than middle-class female speakers. We expect that the relative extent of usage will be highly variable, and depend both on the speaker and the situation, as social gender is highly context dependent. But the overall pattern should be that the working-class speakers make greater use of engendered variation that middle-class speakers. This is reflected in Figure 4.1 by the larger sex differentiation among the working-class than the middle class for both variants.

One of the characteristics of phonetic-level engendered variation is that it shows socially-determined exceptions to the general pattern that women use more feminine variants, and men use more masculine variants. As we have already seen in Stuart-Smith’s study of sibilant centre of gravity in Glasgow (Figure 2.4), where the young working-class female teenagers produced the masculine variant (low frequency sibilant

**Social Class Hypothesis 2:** Working-class speakers show greater sex differentiation for both masculine and feminine engendered variants than middle-class speakers.
centre of gravity), there are exceptions to the Social Class Hypothesis. In such exceptional cases, we need to further consider factors other than social class, of which the most important is sure to be social gender itself.

4.2.2 An example of nongendered variation and social class

In Chapter One, I claimed that nasality was a nongendered source of phonetic-level variation because men and women do not show dimorphism in the anatomy of the velum. I listed several studies that failed to find aerodynamic and acoustic differences and suggested that there was no reason why we might expect one sex to use a nasal voice quality more often than the other. The fact that nasal voice quality in a nongendered source of phonetic-level variation has implications. Namely, we predict that since the nasal voice quality is nongendered, we will not see any of the patterns I am claiming engendered phonetic-level variation shows. A number of studies on nasality have shown that nasal voice quality varies systematically by sex and social class. However, we are not concerned with idiosyncratic variation between and within speech communities. Rather, we are looking for recurring patterns that show up again and again from one study to the next. With this in mind, let us now take a look at the results of these studies.

Esling (1978) interviewed 52 male adults and 8-year-old boys from various socioeconomic backgrounds in Edinburgh, Scotland. Each speaker was assigned to one
of three status groups, high, middle, or low, depending on their occupation, type of housing, area of residence, and educational background. The children were assigned to the same status as their fathers. Esling carried out an impressionistic analysis of 34 laryngeal and supralaryngeal voice quality features, including nasal voice. Esling found that the higher the status group, the more often a voice was classified as ‘nasal.’

In another study of voice quality, this time in Vancouver, Esling (1991) again found that the highest status group that he looked at had the most nasality in their voice. He interviewed 64 adults between the ages of 16 and 35, divided evenly between the sexes and four socioeconomic groups ranging from middle working class up to middle middle class. He found that the middle middle class pronunciation of vowels was characterized by nasalization (p.126).

On the other hand, Trudgill (1974:187) reports that “an auditory quality which can be loosely labelled as nasality is a frequent component of the Norwich working-class setting.” It seems that there is a lack of a real trend in the association of a nasal voice quality with a specific social class.

Stuart-Smith (1999), in her study of voice quality discussed in §2.6.2, reported that greater nasalization was found in the speech of the male speakers. Although social class was considered in the study, she did not note any class differences in the use of nasality (p.218).

Finally, consider nasality and politeness. It has been reported that nasal voice quality can also be used as a marker of politeness, a characteristic of feminine speech. Crystal (1975:87, referring to Key 1967:19), writes:

In Cayuvava, a rapidly-disappearing language of Bolivia... an individual of lower social or economic status addresses one of higher rank with a prominence of nasalization for all vowels of the utterance; and similarly with a women being polite to her husband, or a man asking a favour.

This use of a nasal voice quality as a marker of politeness in Cayuvava is similar to the use of breathiness and average pitch as markers in politeness in Japanese (see §2.4.3 and §2.5.2). Recall that since high pitch correlates with the stereotype of physical diminutivity, a high-pitched voice is considered to be non-aggressive, and therefore polite (Loveday 1981; Ohala 1983, 1984; Van Bezooijen 1995).
The difference between a nasal voice and a high-pitched voice is that the nasality is not inherently associated with feminine speech. Because nasality is not engendered variation, we predict that we will not see an overall pattern, and that is exactly what the five studies reviewed here show. The results obtained by Stuart-Smith in Glasgow and Trudgill in Norwich suggest that nasality is associated with masculinity and working-class speech. However, the results obtained by Key (reported in Crystal) and Esling suggest that a nasal voice quality is associated with middle-class speech and politeness. This inconsistency of the way nasal voice quality patterns with social class contrasts sharply with the consistency of patterns of engendered variation, and adds further weight to the argument for a distinction between the engendered and nongendered phonetic-level variation as laid out in Figure 1.1. In the case of engendered variation, we predict that we will see consistent patterns with social class, and, at least in the data reviewed here, that is what we see. In the case of nongendered variation, there is no reason for us to suspect that it will pattern with social class, and we do not find a pattern between social class and nasal voice quality.

4.3 Engendered variation, speech styles, and gender theory

4.3.1 Why the “prestige” model fails us

Traditionally, “speech style” has been a sociolinguistic cover term among Labovian sociolinguists for amount of self-monitoring a person does while speaking, and shifting between speech styles was thought to account for much of the intra-speaker variation heard during the interview (see for example, Labov 1972:79 ff.). Speech styles ranged from casual, involving the least amount of self-monitoring, through to minimal word pairs, which focused the speaker’s attention on subtle differences in the pronunciation of similar words such as caught and cot, and therefore involved the most amount of self-monitoring. Speech style correlates with the frequency of the use of vernacular forms, so that casual speech produced the greatest number of vernacular forms, and the reading of minimal pairs the fewest vernacular forms.

In the previous section, I discussed the link between masculine speech and working-class speech. Working-class speech is also associated with vernacular speech, since working-class speakers as a group tend to produce more vernacular forms than
middle-class speakers. By inference we might begin this section by positing a link between masculine speech and casual speech. The idea would be that when a person speaks in a casual tone of voice, they tend to use more masculine speech. As their speech becomes more formal, or self-monitored, they gradually switch to a style of speaking that is more feminine. By masculine speech and feminine speech, I am referring to phonetic-level engendered variation. So for example, a creaky voice might be considered casual, and correlate with other indicators of casual speech. In the same manner a breathy voice would be considered more formal.

We would expect such a link if we consider any graph of sociolinguistic variation by speech style and social class, such as Figure 4.2. This figure shows that class and style complement each other so that the higher the social class and the more monitored the speech style, the more often the standard variant is used (as reflected by the low index score). Figure 4.2 is from Labov’s (1966) study of r-dropping in New York City. Labov counted the frequency of two variants of (r), a fully realized alveolar approximant /r/, and Ø, or r-dropping, which is normally the substitution of a schwa for the approximant. The first form is the standard variant in North American English, while the second variant is a salient part of the New York accent. Because there are only two variants, the index scores for (r) are simply the percentage use of the schwa variant. Figure 4.2 shows that the higher the social class and the more monitored the speech style, the more often (r) is realized as an alveolar approximant, or the standard variant.

The correlation of standard language use with social class and speech style as shown in Figure 4.2 seems to be ubiquitous with a few noted exceptions such as hypercorrection. Given this well-established link between working class, vernacular forms, and speech style, and the link between working-class speech and masculinity claimed by researchers such as Labov and Trudgill, it only seems natural that we should be able to connect these two links to form a chain connecting masculine phonetic-level engendered variation with vernacular speech and casual speech style. The flip side of this hypothesis is that feminine speech is connected to standard speech and a more formal speech style. To continue the example of creak and breathiness, this predicts that a factory worker might use a creaky voice when chatting with his workmates during a smoke break, but a more breathy voice when reporting to the floor manager.
Unfortunately, phoneticians do not normally consider the factor of speech style when they set up their experiments. Rather, the accepted practice is to control for the style of speech by always using a prepared text for the subject to read. As such, there is very little evidence that we can use. Of all of the different studies reviewed in Chapter Two, only one researcher, Stuart-Smith (1999), reported speech style results. A comparison of her table of voice quality descriptions for conversational speech with her table of voice quality descriptions for word lists reveals that overall the differences between the two speech styles for breathy voice and creaky voice (reported in §2.6.2) by speech style are almost non-existent. For the creaky voice setting, the following groups show a slight increase as we move from casual speech to word lists: older middle-class males, older working-class males, younger working females, and younger middle-class females. The following groups show a slight decrease: younger middle-class males and older working-class females. The last two groups, younger working-class males and older middle-class females, do not show a change in the amount of creak. For the breathy setting, only the younger middle-class females showed a slight increase in the use of breathy voice as we shift from casual speech to word list speech. The older middle-class females showed a slight decrease, while the younger working-class females showed a

Figure 4.2  An example of variation of (r) by speech style and social class. As the speech style becomes more formal, and the social class higher, the use of (r) becomes more standard. From Labov 1966, Figure 11. Data approximated from his figure. Following Chambers (2003:24) the values are reversed so that the standard variant is zero.
noticeable decrease. All other groups do not show a difference in the use of the breathy voice quality between the two speech styles. So not only do we not see the predicted pattern, but we do not seem to see any real pattern at all.

Why then, do we get such robust differences between speech styles for variables such as Labov’s r-dropping while Stuart-Smith’s voice quality descriptions do not show any clear pattern? It is not because r-dropping is phonological variation whereas voice quality is a phonetic variation; work by Labov (for example 2001b) and others has shown that phonetic-level variables also behave in same manner as the (r) variable. But there are two important differences. The first is that r-dropping seems to be a part of a change in progress, whereas, while Stuart-Smith does discuss voice qualities changes between the older and the younger generation, the use of breathy and creaky voice seem to be stable. However, engendered variation certainly can and does participate in language change (see §4.4).

More importantly, the r-dropping and the other phonetic-level variables such as vowel fronting that show the same pattern as seen in Figure 4.2 are nongendered variation, whereas the creaky and breathy voice settings are engendered variation. To understand the significance of this difference, we must first go back to Figure 4.2 and try to understand the forces behind the variation. Labov (1972:23), discussing an ongoing change of diphthong centralization on Martha’s Vineyard, writes:

At the first stage of change, where linguistic changes originate, we may observe many sporadic side-effects of articulatory processes which have no linguistic meaning: no socially determined significance attached to them… Only when social meaning is assigned to such variations will they be imitated and begin to play a role in language.

The social meaning that Labov is referring to is “prestige” (p.3). Labov, in his “prestige model” of sound change (p.178–80), claims that a linguistic variant gains positive or negative prestige by becoming a marker of the speech of different social class groups. Trudgill (1972, previously discussed in §4.2.1) further elaborated on this model by claiming that the “negative prestige” associated with the speech of the working class is actually positive prestige, what he calls “covert prestige,” to men.

It is certainly possible that an engendered variant such as a fully voiced /h/ could become associated with the speech of the working class, and therefore become
stigmatized. However, unlike variables such as (r) in New York City or diphthongs in Martha’s Vineyard, to which the only social significance that can be associated with them is prestige, engendered phonetic-level variation has another type of social significance inherently associated with it: gender. Regardless of whether or not a fully voiced /h/ is considered stigmatized because it is associated with the speech of the working class, it will most likely have a masculine connotation. Engendered variation will only show behaviour similar to that seen in Figure 4.2 if there is prestige social significance associated with it.

If on the other hand, as seems to be the case for breathiness and creak in Glasgow, the variation has only gender social significance, then examining speech styles such as casual speech and word list is not appropriate; without the prestige element, there is no longer a reason to believe that engendered variation will vary systematically by speech style. Thus, Labov’s prestige-oriented speech style framework is not the best with which to examine engendered phonetic-level variation. Of course, we will always need an examination of speech styles to determine if there is prestige social significance associated with the variation.

The claim that prestige-oriented speech styles are not an appropriate framework for the study of phonetic-level variation is not new. Argente (1992) contrasts “ways of speaking” with “speech styles,” arguing that they are not the same. To her, speech style is connected to the amount of self-monitoring, whereas way of speaking is concerned with specific speech events, such as issuing a warning, praying, whispering, appealing, etc. (but c.f. Hymes’s (1972) discussion of sociolinguistic competence). A speech style is determined by the frequency of several linguistic variables acting together. Argente compares the sociolinguist’s approach (the concern for frequency of segmental occurrence) with that of the phonetician, and their concern with suprasegmental variation (vocal effort, co-articulation, breathiness, etc.). Building on the speech style versus way of speaking distinction, she suggests that sociolinguists and phoneticians are dealing with two distinct phenomena in stylistic variation. The phonetician’s stylistic variation is intralocutor (i.e., situational) variation, whereas the sociolinguist’s stylistic variation is interlocutor (i.e., intergroup) variation. She concludes that there are two types of variation: speech style, which marks group membership within the community, and way
of speaking, or speech variation expressing a speaker’s response to distinct socially-defined situations within the community. However, Argente’s claim is too strong; it denies the possibility that nongendered phonetic-level variants can have prestige associated with them. I do not see why phonetic-level variables cannot vary in both speech style and speech situation.

4.3.2 Gender Theory and gender styles

Modern gender theory claims that gender is not something static, that is, an inherent and invariable trait, but rather something dynamic that we enact (Eckert and McConnell-Ginet 1992). Not only our social background such as our age and social class, but more importantly, the networks that we belong to, help set the community standards by which we then respond to and build upon in a gender-appropriate way. Eckert and McConnell-Ginet cite the example of the stereotype of women as tending to speak in an indirect manner (p.471). This stereotype is based on the mothering practices of white middle-class women, and they cite Tannen (1982) as evidence that it does not apply very well to African-American women. Eckert and McConnell-Ginet call these groups such as middle-class white women “communities of practice,” and claim that we cannot discuss gender without first understanding the “community of practice” by which gender is judged: “to interpret broad sex patterns in language use without considering other aspects of social identity and relations is to paint with one eye closed” (p.471). They caution:

What many of the studies [of gender differences in language use] cited above have found are tendencies towards gender-differentiated practice that have implications for language. It is important to remember that statements like “women emphasize connection in their talk whereas men seek status” are statistical generalizations… Analysts all too often slide from statistical generalizations to quasi-definitional or prototypical characterizations of “women” and “men,” thus inaccurately homogenizing both categories and marginalizing those who do not match the prototypes. (p.470, original emphasis)

Throughout the first four chapters of this thesis I discussed several tendencies of gender differentiation at the phonetic level. In every case, I compare “men” to “women,” while playing down the role of the individual, and therefore running the risk of “inaccurately homogenizing both categories and marginalizing those who do not match the prototypes.” To avoid this, Eckert and McConnell-Ginet recommend that we:

abandon several assumptions common in gender and language studies: that
gender can be isolated from other aspects of social identity and relations: that gender has the same meaning across communities, and that linguistic manifestations of that meaning are also the same across communities. (p.462)

While the abandonment of the first two assumptions is certainly a positive step forward with respect to Gender Theory, that the third assumption holds for phonetic-level engendered variation is one of the tenets that the engendered phonetic-level variation framework is built on. The social significance of speaking in an indirect manner may vary across communities, but if we abandon the assumption that phonetic-level engendered variation has the same meaning from one community to the next, then the engendered variation enterprise is for naught. Fortunately, there is evidence such as Ohala’s frequency code cited in §4.2.1 that the manifestations of gender based on sex differences in anatomy of the vocal tract have the same meaning across all communities.

So according to researchers such as Eckert and McConnell-Ginet, gender is constructed and performed. We can then speak of “gender styles” such as masculine style and feminine style. Furthermore, since social gender is not discrete, there are different degrees of masculine and feminine styled speech. I will refer to these styles in terms of strength, so that, for example, to speak in a style that makes use of a large number of masculine variables is to speak in a strong masculine style.

If gender is something that is completely constructed, then engendered phonetic-level variation should also participate in this construction. In Chapter Two, I mentioned a California Latina gang girl’s use of average pitch (§2.4.1). She speaks in a high-pitched voice to her boyfriend and mother, but in a low-pitched, creaky voice to other female gang members. This is one example of the construction of gender via phonetic-level variation. Following are two more, the first from a woman working in what is regarded as traditionally a man’s job, and the second from a man posing as a women in order to do a woman’s job.

The first example of gender performance is taken from Bonnie McElhinny’s (1995) study of women in the police force. The following is an excerpt from an interview with Janie, a European American rookie:

Bonnie: Do you think women on this job start to act in masculine ways?
Bonnie: Like what are some of the things you see?

Janie: Your language. I know mine changes a lot. When I’m at work I always feel like I have to be so gruff, you know. And normally I’m not like that… Mine’s atrocious sometimes. I’ve toned it down a lot. When I first started you know cause I worked with a lot of guys it seemed like- They didn’t- may not even have swore but I felt like I had to almost like be tough or something around them….

Bonnie: Is it mostly profanity, or do you do it with tone of voice or something?

Janie: Little bit of both…. Like I said about how Black women were able to kinda command respect from people in the projects. I try to like pick up some of their slang, either their slang or their tone, something.

Unfortunately Janie does not use any words that refer specifically to phonetic-level masculine variation per se, but the implication is that she is trying to use a masculine “tone of voice” along with a large amount of profanity to sound “tough.” Notice the contrast between Janie’s description of her speech at work and her speech elsewhere: Janie is adopting a strong masculine style at work that she does not normally use.

The second example of the enactment of gender, a sex fantasy line telephone worker (Hall 1995), takes the acting part literally. Andy is a thirty-three-year-old Mexican American bisexual who portrays himself as a female heterosexual with his male callers. Andy, in his interview with Kira Hall, reports that he is best received when he performs a stereotypical feminine voice (p.202–3):

… And the other thing I’ve found over the years is it’s better to sound soft and quiet than loud and noisy … if you’re a women. … [It’s] better to sound ((whispered)) soft you know, softer. ((in a natural voice)) You know, like whispering, rather than ((in a loud voice)) OH HO HO HO, ((in a natural voice)) really loud, you know…

---

7 Hall uses the following transcription conventions (.211):

(2.0) indicates length of pause between utterances in seconds

((   )) double parentheses enclose nonverbal movements and extralinguistic commentary

[   ] brackets enclose words added to clarify the meaning of the text

__ underlining indicates syllabic stress

CAPS upper case indicates louder or shouted talk

: a colon indicates the lengthening of a sound

. a period indicates falling intonation

, a comma indicates continuing intonation

? a question mark indicates rising intonation

^a^ rising arrows indicate higher pitch for the enclosed segment

… deletion of some portion of the original text
So here, I’ll give you the voice, okay? Hold on. (4.0) ((in high pitch, soft whisper)) Hello. (2.0) Hello? (2.0) How are you? (1.5) This is Emily. (in natural voice) See? … ((inhales, then in slow, high, breathy voice)) Hello::: Hi::: (gasps) Oh yes! (0.5) I’m so horny right ^no^::w…((in natural voice)) It’s funny how I’ve actually taped myself and then played it back, and it’s actually two separate voices.

Andy makes liberal use of a strong feminine style to overcome and disguise the fact that he is a man portraying a woman. In the above quotation, there are at least three feminine variants: high-pitched voice, breathy voice, and lengthened vowels. Although Andy’s speech is in some respects artificial because of its performative nature, his description of the way he switches back and forth between his natural voice and the soft, high-pitched, breathy voice of Emily is a good example of gender construction.

These examples show that an individual is able to fine tune the strength of their phonetic-level engendered variation according to the situation at hand. So gender styles are a better framework for the examination of phonetic-level engendered variation than speech styles. Following researchers such as Eckert, McConnell-Ginet, and McElhinny, we need to solicit different gender styles either from the same individual (as seen in the conversation between Hall and the Mexican American bisexual), or from different individuals, by comparing and contrasting, for example, the casual speech of female police officers with the casual speech of female hairdressers.

The discussion of gender speech style in this chapter has been based solely on the research of linguistic anthropologists. These researchers look at many types of language variation, ranging from pauses and interruptions to body language, but they seldom look at the phonetic-level variation discussed in chapters 2 and 3. The linguistic anthropological research tends to lead to the same conclusion: gender is constructed to fit the speech situation at hand. But it is not so clear that we can make the same claim about engendered phonetic-level variation. How much of engendered phonetic-level variation really is dynamic and how much is static is a question that has not been adequately addressed by anthropologists, sociolinguists, or phoneticians.

A number of phoneticians have suggested that the engendered variation discussed in Chapter Two that falls under the label “voice quality,” such as breathy voice, creaky voice, etc., is a semi-permanent component of our voice:

The term ‘voice quality’ refers to those characteristics which are present
more or less all of the time that a person is talking: it is a quasi-permanent quality running through all the sounds that issue from his mouth. (Abercrombie 1967:91)

Voice quality is conceived here in a broad sense, as the characteristic auditory colouring of an individual speaker’s voice... Perceptually, voice quality in this broad interpretation is a cumulative abstraction over a period of time of a speaker-characterizing quality, which is gathered from the momentary and spasmodic fluctuations of short-term articulations used by the speaker for linguistic and paralinguistic communication. (Laver 1980:1)

The term *voice quality* refers to those features of speech present more or less all of the time that a person is speaking — the background characteristics perceived as the most constant or persistent over time (Esling 2000:26)

These comments suggest that at least some of the phonetic-level variation we have been discussing is not readily malleable, at least in the short-term.

Furthermore, recently the term voice quality has been applied not only to suprasegmental variation, but to segmental variation as well. For example, Biemans (2000) gives the following example (following Laver and Trudgill 1979 she uses the term “setting”):

The span of a setting can range from short-term, via medium-term, to long-term. A short-term, linguistic example of a setting is co-articulation, when, for example, the rounded properties of a vowel (e.g. [o]) are transferred to a consonant preceding it. (p.21)

By lumping in the short-term articulation of segments with long-term voice quality, we can speak of, for example, the mean sibilant centre of gravity (segmental quality) and breathiness (suprasegmental quality) in the same terms of permanency. A “permanent” segmental setting entails that the speaker articulates that segment consistently in the same manner regardless of speech or gender style. For example, it may well be that the low mean centre of gravity for /s/ for the working-class teenage girls from Glasgow (see §2.3.1) is a permanent setting. If so, sibilant centre of gravity will not participate in the active construction of dynamic gender. At the same time, these passive gender elements are always there in the background to be picked up on by listeners, and therefore they still play a role in the construction of gender. If there really is such a distinction between temporary, situational markers of gender and semi-permanent markers of gender, then we need to make a distinction between *dynamic* gender and *static* gender. Dynamic gender is created and adjusted depending on the situation, such as that Latino gang girl’s use of
pitch. Static gender does not vary by situation, but rather is pretty much always present in
the speech signal.

But how do we go about testing such claims about the dynamic or static nature of
phonetic-level engendered variation with respect to the construction of gender? One
possibility is to posit a “gender awareness” that parallels Labov’s “social awareness.”
Gender awareness tells us what is appropriate engendered speech and what is
inappropriate engendered speech. We must have an awareness of this nature in order to
pick up on and act out gender in our “community of practice.” If so, then this naturally
raises the question: Is there variation below the level of gender awareness, much like
there is variation below the level of social awareness? The existence of such an
awareness is our first hypothesis on gender styles and phonetic-level variation.
Engendered variation below the level of gender awareness would then be a “permanent”
voice quality setting. This approach allows us to fall back on the already well-established
speech perception work done by Labov and others on linguistic variation and social
awareness.

**Gender Style Hypothesis 1:** There is engendered
variation both above and below the level of gender
awareness. Phonetic-level engendered variation
below the level of gender awareness does not
participate in the dynamic construction of gender.

### 4.3.3 “Tomboys” and “sissies”

Masculine behaviour in women is often less stigmatized than feminine behaviour in men
(Eckert and McConnell-Ginet 2003:37). The stigmatization of feminine behaviour among
men begins in early primary school (Thorne 1990:110):

Boys who frequently seek access to predominantly female groups and
activities (“sissies”) are more often harassed and teased by both boys and
girls. But girls who frequently play with boys (“tomboys”) are much less
often stigmatized, and they continue to maintain ties with girls, a probable
reason that, especially in the later years of elementary school, crossing [into
the other gender’s social networks] by girls is far more frequent than
crossing by boys.
I assume that this asymmetry in the stigmatization of masculine and feminine behaviour will be reflected at the phonetic level. This motivates our second hypothesis on gender styles and phonetic-level variation:

**Gender Style Hypothesis 2:** Women use masculine variation that is above the level of gender awareness more often than men use feminine variation that is above the level of gender awareness.

The *Gender Style Hypotheses* are not empirically tested in this dissertation. They are presented here because they further demonstrate the importance of the nongendered–engendered variation dichotomy, the backbone of the engendered variation framework.

### 4.4 Engendered variation and language change

#### 4.4.1 Men, women, and language change

Labov (1990) introduces three principles concerning the role sex plays in language variation and change (wording from Labov 2001a):

**Principle I** For stable sociolinguistic variables, women show a lower rate of stigmatized variants, and a higher rate of prestige variants than men (p.266)

**Principle Ia** In linguistic changes from above, women adopt prestige forms at a higher rate than men (p.274)

**Principle II** In linguistic change from below, women use higher frequencies of innovative forms than men do (p.292)

These principles are concerned with the three types of language variation: stable stratification, changes from below, and changes from above. The latter two are with reference to the level of social awareness (Labov 1966:328). Changes from below the level of social awareness are motivated by internal linguistic factors, and do not have prestige associated with them. They tend to go unnoticed by the community until the change is near completion (Labov 1994:78). Changes from below, because they are
below the level of conscious awareness of the speech community, are difficult to detect, and show little correlation with speech style (Labov 1966:328; 2001a:196).

Changes from above the level of social consciousness involve linguistic variables that have prestige associated with them. Changes from above either involve the importation of a new prestige feature from outside the community, or the re-distribution of a prestige form within the community. Such changes show a strong correlation with speech style, with a higher rate of occurrence of the prestigious variant in the more formal styles of speech. As well, changes from above are subject to hypercorrection and sometimes participate in the formation of overt stereotypes (Labov 2001a:272–4).

In Labov’s writings, the only type of social significance is that associated with prestige. As I pointed out in §4.3, I envision engendered variation as being associated first and foremost with gender social significance, and may or may not be associated with prestige. If a phonetic-level engendered variable is associated with prestige (and at this point it is not even clear if this a possibility), then I assume that it will behave like any other linguistic variable with the same level and type of prestige. We are, therefore, only concerned with variables, both stable and those changing, that act below the level of consciousness. In other words, I assume that prestige social significance will take precedence over gender social significance in cases where both are applicable.

With regard to stable phonetic-level engendered variation, the wording of Principle Ia is inappropriate for two reasons. Firstly, the principle assumes that all stable variation will always have prestige associated with it. At the phonetic level, this is of course not the case. The women in Spain, England, Japan, and the United States (see §2.5) presumably use a breathier voice quality than men because it makes them sound feminine, not because a breathy voice quality is more prestigious in these areas of the world. In fact, Henton and Bladon (1985) found that women who spoke with a RP accent, the prestige dialect of England, were less breathy than women who spoke with a Northern accent (as mentioned in §2.5.2). Secondly, the principle explicitly states which sex will use more of a variable. But if the variable under discussion is masculine variation, then by its nature men will produce more of it. If, on the other hand, we are talking about feminine variation, then we predict that women will produce more of it.

The second point also applies to Principle II: we cannot state which sex will use
more of the engendered variable without first stating if it is a masculine or a feminine variable. These claims are summarized in the following two Language Variation and Change (LV&C) Hypotheses:

**LV&C Hypothesis 1a:** Men use a masculine variant more frequently than women regardless of whether the variation is stable sociolinguistic stratification or a change-in-progress, as long as it is void of prestige.

**LV&C Hypothesis 1b:** Women use a feminine variant more frequently than men regardless of whether the variation is stable sociolinguistic stratification or a change-in-progress, as long as it is void of prestige.

As an example of *LV&C Hypothesis 1*, consider the merger of [ə] and [a] (the COT–CAUGHT) in Pottsville, Pennsylvania (Herold 1990, as reported in Labov 1994:317–22). In Pennsylvania, as in some other areas of North America, the low back vowels [ə] and [a] are merging. In 1977, members of the Project on Language Change and Variation (see Labov 1994:xi) carried out a sociolinguistic survey of the merger in a number of cities and towns in Eastern Pennsylvania, including 13 youth, aged 12 to 14 years, who live in Pottsville. In 1988, Herold returned to Pottsville and obtained minimal pair results for 16 youth of the same age as the first study. Figure 4.3 shows the percent of tokens that were merged during each study by sex. Labov claims that the results of these studies show a feature that is characteristic of mergers: “that females are in advance of males” (Labov 1994:319). However, that is clearly not the case for the earlier study, with the boys leading the girls by eight percent. It is not until the second survey that the girls lead the boys, and then it is by a dramatic amount. But by the second survey, the merger is approaching completion for the girls. This sudden and dramatic change in which sex is leading the merger in Figure 4.3 is the intersection of following points I have made in this section and elsewhere in the dissertation:

- a merger results in the loss of phonetic distinctiveness
the loss of phonetic distinctiveness is masculine variation (Chapter Three)

therefore men lead women in mergers as long as they are void of prestige
(LV&C Hypothesis 1)

changes from below are void of prestige (this section)

a change from below can rise to above the level of social consciousness as it
nears completion (this section)

women lead changes from above (this section)

If between the years of 1977 and 1988, the merger of the low back vowels in Pottsville
rose to above the level of social consciousness for the girls as it neared completion, then
we have a complete account of the linguistic behaviour of the teenagers seen in Figure
4.3. The boys initially lead the merger because they are male and it is masculine variation.
The girls lead the merger in its final stages because the merger has reached the point
where it is above the level of social consciousness, and women lead changes from above.
In Chapter Six, I further challenge this claim that “females are in advance of males” with
data from thirteen mergers taking place in North American English.

4.4.2 Language change and anatomical engendered variation

In this section and the following section, I introduce data from apparent time studies of
phonetic-level language change. The objective of these sections is twofold. The first
objective is to provide more examples of language change from the perspective of the
phonetic-level engendered variation framework. The second objective is continue to posit
hypotheses about language change and phonetic-level engendered variation. In this
section, I strengthen the LV&C Hypotheses with data from an ongoing language change
in British English. We will see examples of both feminine engendered variation and
masculine engendered variation. As expected, the women are leading the change towards
the feminine engendered variant, while the men are leading the change towards the
masculine engendered variant.

The voiceless alveolar plosive /t/ in word-medial and word-final position is
currently under going change in British English. In both positions, the plosive displays a
large range of variants, including the following (Docherty et al. 1997:293; Mathisen
1999:115):

i. [t]  – fully released voiceless aspirated / non-aspirated plosive

ii. [ʔ]  – glottal stop

iii. [ʔt]  – double articulated glottal stop and alveolar plosive

iv. [ɾ], [t]  – voiced alveolar tap, or a variant that sounds like [d]

v. [ɹ]  – voiced alveolar approximant

Early commentators on the change suggested that the standard variant (variant i.) in mainstream middle-class English is gradually being replaced by the glottal stop (variant ii.), a borrowing from the speech of the urban working class (for example, Wells 1982:106). Wells made a distinction between “t-glottaling,” the replacement of the alveolar plosive with a glottal stop (variant ii.), and “glottal reinforcement,” a simultaneous double articulation of the plosive and the glottal stop (variant iii.). Milroy et al. (1994:6) emphasize that this distinction between the two glottal variants is important since not only do they occur in different environments, but they also pattern in opposite ways with respect to class and sex in the dialects of England. Milroy et al. began

Figure 4.3  Progress of the COT–CAUGHT merger in Pottsville, Pennsylvania. From Herold 1990, as reported in Labov 1994, Figure 11.4. Values are an average of Figure 11.4a and Figure 11.4b.
their argument by citing data from Mees (1987), a study of (t) variation in the speech of 36 young teenagers from three social class groups in Cardiff. Mees found that the girls produced the standard variant (i.) and glottal stops (ii.) more than the boys, while the boys produced the voiced tap (iv.) and the glottalized variant (iii.) more than girls. This was followed by data from Rigg (1987), an investigation of the glottalization of all three voiceless plosives /p t k/ in Tyneside. The same gender pattern as in Cardiff — men produce more glottalized plosives than women — was found in Tyneside. Figure 4.4 shows the distribution for /t/. Milroy et al. explain the different patterning by sex of the glottal stop and the glottalized plosives as follows (p.25–6): the glottalized variants are well-established features of the local dialect, whereas the t-glottaling is a recent adoption. Furthermore, t-glottaling is associated with the RP accent, and therefore has the corresponding social significance of prestige associated with it. We then see that the sex differentiation pattern seen in these two communities agrees with Labov’s Principle Ia — in linguistic changes from above, women adopt prestige forms (in this case the glottal stop) at a higher rate than men.

Recent work by Docherty and Foulkes (Docherty and Foulkes 1999, Foulkes and Docherty 2006) suggests that the use of the glottalized variant may be increasing over time in Newcastle. They examined variation in word-final and word-medial stops between sonorants in word list productions and conversations of 32 Newcastle residents and 32 Derby residents. Each group consists of an equal number of men and women, younger (15 to 27 years) and older (45 to 67 years) speakers, and working class and middle class. The Newcastle data for word-medial position is of particular relevance. Docherty and Foulkes report that unlike the RP variety of English, Newcastle English shows an almost complete lack of the glottal stop variant (ii.) in word-medial position between sonorants, preferring instead either the plain stop (i.) or the glottalized variant (iii.). Docherty and Foulkes switched the label from “glottalized” to “laryngealized” in their most recent article, and describe the variant as “a lenited stop with laryngealized or creaky phonation” (Foulkes and Docherty 2006:413). In chapters Two and Three I suggested that creak and lenition were masculine phonetic-level engendered variants. As such, we expect that the lenited / creaky variant will be associated more with male speech than with female speech, and that is exactly what Docherty and Foulkes found: the males consistently used more of the laryngealized variant than their age and class female
counterparts. Furthermore, the younger speakers show a marked increase in the use of the laryngealized variant, with the working-class young males showing the greatest use of that variant. These data suggest that the laryngealized variant is increasing in Newcastle English, and that the males are leading the change.

But when we turn to the word-final /t/, a completely new set of patterns emerges for Newcastle. Again, unlike the RP accent, the glottal stop variant of /t/ in Newcastle English is almost non-existent in pre-pausal position. What we find instead are the following variants of released /t/:

i. \([t^h]\) – a voiceless alveolar plosive followed by a release burst and a short period of aspiration

ii. \([t^h]\) – a voiced alveolar plosive followed by a release burst

iii. \([t^{h^h}]\) – a voiceless alveolar plosive preceded by a brief period of frication and followed by a release burst

Surprisingly, the canonical form (i.) was the least frequent variant in both dialects, accounting for 24 percent of the Derby tokens and only nine percent of the Newcastle tokens. The other two variants also showed a remarkable difference by region. The
Derby residents do not produce the “extended frication” variant (iii.) at all, and the “continued voicing” variant (ii.) only occasionally. The Newcastle residents, on the other hand, make liberal use of both. Figure 4.5 shows the break down by sex and age group for the Newcastle residents. The continued voicing variant shows little variation by age for word lists, but elsewhere Docherty et al. (1997:Table 4, reproduced here as Table 4.2) report that the voiced variants in casual speech for the same corpus are on the rise. For both age groups and both speech styles (word lists in Figure 4.5), the men produce this variant more often than the women. Recalling from §2.2.2 (particularly the discussion of Koenig 2000) that males tend to voice more often than women, we see that this is yet another example of a masculine engendered variant being used more frequently by men.

As well, from the same figure we see that the younger speakers produced the extended frication variant notably more often than the older speakers, suggesting that it is a change in progress, with the use of extended frication on the rise in Newcastle. For both age groups, women produced this variant more often than men. Again, if you recall from §2.2.2 (also particularly the discussion of Koenig 2000) and §2.5.1 that women tend to devoice / produce breathy voice more often than men, then we see that this is the example of feminine engendered variation that we are looking for. Although the
extended frication variant is a highly localized, non-standard variant (parenthetically, women in both regions use more of the “standard” variant than men) as shown by its lack of use in Derby, women use it more often than men.

So now we have seen two examples of phonetic-level engendered variation, one masculine and one feminine, that are both changes in progress. In the case of the masculine variant, men are showing an increased usage over time, whereas in the case of the feminine variant, women are showing an increased usage over time. From this, it should be clear that Labov’s Principle II — in linguistic changes from below, women use higher frequencies of innovative forms than men do — is not applicable to engendered variation. The problem with Labov’s principle is that it makes a claim about the behaviour of the sexes. This claim is in conflict with my claim that men use more masculine variants and women use more feminine variants, as long as the variation is below the level of social consciousness (LV&C Hypothesis 1a and 1b). It seems that when prestige is not a factor, then the only social significance phonetic-level engendered variation has is social gender, and it this that determines the linguistic behaviour of men and women.

4.4.3 Language change and behavioural engendered variation

The main claim of Chapter Three was that women produce sharper phonetic-level distinctions than men. This was seen in the voiced / voiceless contrast for plosives (§2.2.2), vowel space (§3.2), phonemic length contrasts (§3.3), and allophonic variation
In all cases, women articulated their speech so that the acoustic correlates of the phonetic contrast were more distinct than those produced by men. In this section, I demonstrate how this behavioural phonetic-level engendered variation impacts on what we know about the way vowels change in North American English. I begin with a simple example that is followed by what is the most important hypothesis of this dissertation. I then use this new hypothesis to re-examine some of Labov’s work on the Northern Cities Shift in light of the phonetic-level engendered variation framework. We also take a look at the Southern Cities Shift, which Labov claims is a counterexample to the generalization that men lead in vowel changes that involve centralization, and therefore is also a counterexample LV&C Hypothesis 1. Lastly, I conclude this section by attempting to resolve what Labov has dubbed the “Gender Paradox,” one of the mysteries of modern sociolinguistic theory.

The example of engendered phonetic-level variation and language change comes from Keating et al. (1994). This paper introduces two phonetic studies carried out on the TIMIT database of American English (see also §3.3). The TIMIT database is a collection of audio recordings of 6300 utterances of read speech by 630 native speakers of American English (each speaker read ten sentences). The speakers are categorized by race, sex, age, and dialect region, with a majority of the subjects being white males in their 20s. The entire corpus is transcribed in phonetic script, and the transcription includes several kinds of subphonemic detail. Keating et al. looked at the pronunciation of the vowel in the word “the” and the effect of vowel context on the acoustics of velar plosives. We will only be concerned with the former study. The TIMIT database transcription system distinguishes between full and reduced vowel qualities. The reduced varieties include schwa [ə] and barred i [ɨ]. The database contains 2202 tokens of the word “the,” the most common word in the database. The vowel quality for the word “the” shows a large amount of variation, of which [ə], [ɨ], and [i] are the most common variants. Furthermore, the following environment plays an important role in determining the variant: the reduced vowels [ə] and [ɨ] tend to occur before a consonant, while the tense vowel [i] tends to occur before a following vowel, as shown in Figure 4.6.

Figure 4.6 shows vowel usage for all of the speakers in the TIMIT database. When the use of the full vowel is compared to the use of a reduced vowel by age group, as is
done in Figure 4.7, an interesting pattern emerges: as the age of the speaker decreases, the use of a reduced vowel increases before vowel.

In §3.3, I discussed vowel reduction and engendered variation. At that time, I introduced a study of vowel reduction by Byrd (1994) on the same database that showed that men used the schwa more often than women. I also introduced a study of West Coast American English (Henton 1990) that showed that men pronounce the lax vowel /ʌ/ as a closer to schwa than women do. Together, these two studies suggest that if we were to break down the data shown in Figure 4.7 by speaker sex, age, and vowel quality, then we should see that young men are leading in the usage of schwa. This assumes, of course, that the change is still below the level of social consciousness.

I chose this example because the variable itself is straightforward, and clearly illustrates how our predictions about the behaviour of engendered phonetic-level variation in general can be extended to make predictions about language change. Specifically, since men produce more reduced vowels than women, and use a more contracted vowel space then women, we predict that men should lead the change from a full vowel to a reduced vowel in the word “the”. Shortly, I discuss two examples that are far more complex, the Northern Cities Shift and the Southern Cities Shift.
Now that we have seen an example of how behavioural engendered phonetic-level variation might interact with language change, we are ready to articulate our next hypothesis. In Chapter Three, we saw that men produce less distinct contrasts between vowels, and reduce to a schwa more frequently. We take this observation as the starting point. If we now add to this that much of phonetic-level language change involves either the maintenance or loss of distinctions (such as vowel categories), then we are ready to articulate our next claim: women lead men in linguistic changes that involves the maintenance of a phonetic distinctiveness, while men lead women in linguistic changes that involves the loss of a phonetic distinctiveness.

**LV&C Hypothesis 2:** Women lead men in linguistic changes that involve the maintenance of phonetic distinctiveness, while men lead women in linguistic change that involves the loss of phonetic distinctiveness, as long as there are not overriding factors (such as prestige) involved.

Although this falls out of the facts of engendered variation and *Gender Hypothesis 1*, I
feel that it is important to articulate this claim as a separate hypothesis because of its importance to sociolinguistic theory in general.

Labov (1990) made an observation about the role of sex in language changes involving vowels that sounds similar to *LV&C Hypothesis 2*:

Some of the first sound changes studied made it seem possible that females led in the upward movement of peripheral tense vowels that *increased the dispersion of the vowel system*… whereas males led in the opposite trend: *shifts that moved towards the center* corresponding to a “closed-mouthed” tendency (p.219, emphasis my own)

A change that increases the dispersion of the vowel system is change that maintains and increases contrasts between vowels. Shifting vowels towards the centre, on the other hand, reduce the amount of contrast between vowels. Labov’s observation about vowel shifts and *LV&C Hypothesis 2* are essentially claiming the same thing: women lead in the maintenance of phonetic-level distinctions. However, I intend *LV&C Hypothesis 2* to apply to all types of phonetic-level distinctions, and not just vowels. Labov goes on to dismiss this generalization because it does not account for consonantal changes led by females such as the devoicing of /ʒ/ in Buenos Aires (Wolf and Jiménez 1979), nor the laxing and centralization of tense vowels before /l/ in the southern United States which leads to a homonymy of minimal pairs such as *steel* and *still*, *sail* and *sell*, *fool* and *full* (Di Paolo 1988). Let us consider these two exceptions to Labov’s generalization before moving on.

According to Labov, the devoicing of /ʒ/ in Buenos Aires (reproduced here as Figure 4.8) is a change from below since there is no overt social reaction to the change, and the variable showed very little variation across speech styles (Labov 1990:217–8). This variable conforms nicely to *LV&C Hypothesis 1b*, that women use a feminine variable more frequently than men in a change in progress, as long as the variable is void of prestige. We saw that the voicing of /h/ (§2.2.2) and the voicing of /t/ (§4.4.2) are masculine variation (the anatomical reasons are discussed in §2.2.2). By corollary, devoicing and breathiness are feminine variants. Thus, I also am not attempting to use the same mechanism to account for both the consonant and the vowel changes, although I am using the same framework of engendered phonetic-level variation. We now have provided a straightforward account of Labov’s first exception: women lead the change of
engendered variation and sociolinguistic theory

Because the change is void of prestige, and because women are more prone to devoicing than men for anatomical reasons. Furthermore, as we have seen from the discussion of the variation of (t) in British English in the previous section, the phonetic-level engendered variation framework is capable of accounting for some of the consonantal variation (namely, the engendered variation), regardless of whether the changes are led by females or males.

The merger of tense and lax vowels before /l/ in the southern United States is not so straightforward. To begin with, further work by Di Paolo (Di Paolo and Faber 1990) on this vowel merger reveals a contradictory story. According to Di Paolo and Faber (1990), the laxing and centralization of tense vowels before /l/ in the southern United States is not a true merger. Although there is no longer a contrast in F1 and F2 between the tense and lax vowel pairs, Di Paolo and Faber found that speakers and listeners make use of other cues to maintain the contrast:

First, it appears that speakers, both in Utah and in New York, make use of differential laryngeal configuration (reflected in our VQI [Voice Quality Index, a measure of amount of creak]) in distinguishing tense from lax vowels…. Second, the great majority of younger Utah speakers no longer have an F1/F2 contrast for most of the tense-lax pairs preceding dark /l/. For them, the laryngeal (that is, VQI) contrast, and perhaps, the

---

Figure 4.8 Percentage /ɔ/ devoiced in Buenos Aires Spanish by sex. Reproduced from Labov 1990: Figure 2. Values estimated from his figure. Original data from Wolf and Jiménez 1979:Table 5.
duration/diphthongization differences may be the only features in their speech available to listeners that distinguish the tense-lax pairs before /l/… Our results suggest that these vowel contrasts are maintained along a dimension other than F1/F2,… (p.201)

Although the formant contrasts are lost between the tense and lax vowel pairs preceding /l/, it seems that the speakers in Utah are making use of other phonetic-level cues to distinguish the two vowel categories, such as voice quality and vowel length.

Furthermore, Di Paolo and Faber found that among the youngest speakers in their survey, there was a trend of reversing the tense and lax vowel categories, so that the tense vowels of the teenagers demonstrated voice quality and formant characteristics of the lax vowels of the older speakers, and lax vowels of the teenagers demonstrated voice quality and formant characteristics of the tense vowels of the older speakers. This led the researchers to conclude that the eventual outcome of this apparent merger will be a reversal in the tense-lax vowels and not a merger (p.199).

This pattern of tense-lax vowel pair reversal is one of the characteristics of the Southern Cities Shift (see for example, Bailey, Wikle, and Sand 1991, Fridland 2000, Fridland 2001, Thomas 2001:Chapter 5). The Southern Cities Shift is considered to be a chain shift, and involves changes in both the high and mid front and back vowels, with the tense and lax vowel nuclei switching places and the back vowels moving forward (Fridland 2001:235). Following are some key points from Fridland’s work on the Southern Vowel Shift in Memphis:

- the changes in the back vowels and the changes in the front vowels seem to belong to two unrelated chain shifts (Fridland 2000:267), with the front vowel shift older than the back vowel shift (Fridland 2000:Figure 5)

- the front vowel changes seem to be occurring above the level of social awareness as shown by the large class differences, are associated with southern rural speech patterns, and are stigmatized by middle-class females (Fridland 2001:241–4)
the back vowel changes seem to be occurring below the level of social awareness, as shown by the lack of general class differences (Fridland 2001:246–9)

the centralization of the back vowels is led by men (Fridland 2000:279; Fridland 2001:246–50)

The back vowel shift displays two of the essential characteristics of masculine phonetic-level variation: it is occurring below the level of social awareness, and it entails the centralization of vowels. As such, our principles predict that men will lead this change, and they do. So in the end, not only does the Southern Vowel Shift not turn out to be a counterexample to Labov’s (and my own) claim that males lead in vowel shifts towards the centre, but on the contrary, appears to be an excellent example. Furthermore, we will see in Chapter Six further data from the Southern Vowel Shift (Labov et al. 2006) that supports this conclusion.

We now turn to the Northern Cities Vowel Shift to consider the role sex plays in the vowel changes in Philadelphia. The Northern Cities Vowel Shift is a chain shift of the

<table>
<thead>
<tr>
<th>Variable and Stage</th>
<th>Female-led</th>
<th>Male-led</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLETED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>backing of (ahr)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>ALMOST COMPLETED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fronting of (æhN), (æhS), (æhD)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>raising of (ahr)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>MID-RANGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fronting of (uwF), (uwC)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>fronting of (owF), (owC)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>NEW AND VIGOROUS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fronting of (aw)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>fronting of (eyC)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>raising of (ay0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCIPIENT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lowering of (i)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>lowering of (æ)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>lowering of (æ)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>raising of (æ)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3 Sound changes in the vowel system of Philadelphia, and which sex leads them. From Labov 1990, Table 3, and Labov 2001, Table 8.5
mid and low vowels occurring in the northern area of the United States around the Great Lakes. Table 4.3 lists the changes and which sex leads each change. Females show more advanced forms in all of the changes but the raising of /ohr/, the centralization of /ʌ/, and the centralization of /ay0/. According to Labov, the centralization of /ʌ/ is linked to the centralization of /ay0/ since they represent parallel shifts of the same nucleus (p.289), while the raising of /ohr/ is a merger with /uhr/ (p.290). Thus we see that women lead the changes that are part of the Northern Cities Vowel Shift and that preserve phonetic contrast, while men lead the changes that involve centralization and the loss of contrast. I provide further empirical support this claim in Chapter Six.

4.4.4 Resolving the "Gender Paradox"

The significance of LV&C Hypothesis 2 becomes apparent once we review the current claims of the role of gender in language change. I began the section on language change and engendered variation with an introduction of Labov’s three principles of the role of sex in language variation and change. Recall that in the case of stable variation and changes from above the level of social awareness, women use higher frequencies of the prestige form than men, while in the case of changes from below the level of social consciousness, women use higher frequencies of the innovative form than men (i.e., they use less of the well-established form, even if it is prestigious). Labov labels this contradictory behaviour of women the Gender Paradox (2001a):

Women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not. (p.292)

Labov calls the juxtaposition of the principles a “paradox” because we are unable to find a “unified explanation of gender differences in both stable and unstable situations,” an explanation we are “impelled” to search for (p.290). Labov (2001a) considers two explanations for the contradictory behaviour of women, both of which he throws out. The first explanation he considers is the possibility that both the conservative and the innovative behaviour of women reflects their superior sensitivity to the social evaluation of language, a claim originally made by Chambers (2003:148–53). In stable situations and changes from above, women are more sensitive to the prestige of the variant and the

---

8 “Superior sensitivity to the social evaluation of language” is Labov’s term. The original wording is “The empirical evidence clearly shows women to be much more able performers than men in the whole spectrum of sociolinguistic situations” (Chambers 2003:147).
stigma associated with the other variants, and therefore react more quickly than men do when it comes to adopting these new variants. In changes from below, women pick up on and act upon the social symbolism that eventually get attached to the new variant (Labov 2001a:291). However, Labov rejects this explanation because “it assigns social sensitivity to early stages of change that are remote from social awareness” (ibid.). This last point is a good one, and worth repeating: any explanation of the role of women as innovators in changes from below that relies on “social awareness” is futile, since changes from below are by definition not associated with prestige. This is also a point we have already dealt with by claiming that even if a change is void of social significance, it can still have gender significance. It is for this reason that we still find women leading changes that are taking place below the level of social awareness in a systematic way.

Not only does the contrast-based account of the role of women in changes from below forge a link with the role of women in changes from above and thus resolve the Gender Paradox, but it also makes concrete, testable predictions. At the beginning of this section I introduced the reduction of the vowel in “the” as a change in progress, and claimed that as long as the change was still below the level of social consciousness, then men should be leading the change since they tend to reduce vowels to a greater extent than women. The implication here is that in general men should be leading changes that result in the loss of phonetic distinctiveness, while women should be leading in changes that result in the preservation of phonetic distinctiveness (LV&C Hypothesis 2). Much of phonetic-level language change involves the maintenance of contrast. The work by Labov and others on vowels forms the foundation of current sociolinguistic theory. Since much of phonetic-level change involves the maintenance of phonetic distinctiveness, we predict that in general women will lead changes from below more often than men. This is, in fact, exactly what we find, as Labov himself has reported: 124

[T]he cases [of sound change] where men are in the lead form a small minority. Furthermore, the male-dominated changes are all relatively isolated shifts. They do not include chain shifts that rotate the sound system as a whole: all such chain shifts examined so far are dominated by women (2001a:284)

This quotation refers to the vowel changes currently in progress in Philadelphia, which are listed in Table 4.3. Not only are the changes led by males fewer, but they are also isolated, and do not play a role in the maintenance of phonetic-level distinctions, exactly
as we would predict from *LV&C Hypothesis 2*. Labov’s comment is based on three decades of research in the way vowels shift, and strongly substantiates *LV&C Hypothesis 2*.

The second explanation of the Gender Paradox that Labov considers is Gordon and Heath’s (1998) sound symbolism proposal. Gordon and Heath note that two important general conclusions of recent sociolinguistic work is that vowel systems tend to rotate in fixed directions (Labov 1994:116), and that females tend to lead linguistic changes (Labov 1990, 2001a). They unite these two observations with the claim that women are attracted to the high front unrounded [i], while men are attracted to the back vowels [a o u] (p.423). This “attraction” is based on what Ohala (1984, 1994) calls the “frequency code,” introduced in part in §4.2.1. Briefly, the high front vowel [i] is perceived as high-pitched, while the back vowels are perceived as low-pitched. High pitch signifies smallness while low pitch signifies largeness. For example, the diminutive in many languages of the world, including English, is marked with [i]. This association between smallness and the high front vowel leads to [i] having “powerful aesthetic value” to women, and therefore women tend to push their vowels in this direction. Gordon and Heath cite many examples form both English and non-English (particularly Arabic) languages in support of their hypothesis. We will examine one of these in more detail here.

In §3.5.2, I introduced a study that looked at allophonic variation of the low vowel when preceded by an emphatic consonant in Arabic. Recall that the central low vowel /a/ is backed to /a/ when preceded by an emphatic consonant. Wahba (1996) looks at this phenomenon from a sociolinguistic perspective. He carried out sociolinguistic interviews with 74 speakers of Alexandria Arabic. During the interview, speakers produced as few as seven and as many as 61 words with low vowels preceded by emphatic consonants. Each token was impressionistically assigned a score based on the following scale:

\[
\begin{align*}
[a:] & \rightarrow 1 \\
[a:] & \rightarrow 2 \\
[æ:] & \rightarrow 3
\end{align*}
\]
Each speaker’s token scores were totalled and then divided by the number of tokens for that speaker to produce an index score of variation in backness. The average index score for each speaker group is shown in Figure 4.9. In general, the educated, older and female speakers have higher average index scores, indicating the use of a more fronted vowel. If we look at gender alone, we see that on average, men use a more backed vowel than women. This is in accordance with Gordon and Heath’s claim that men are attracted to the back vowels, and indeed native speakers report that a greater degree of pharyngealization is perceived as sounding more “masculine” (Kahn 1975:42). The young uneducated women, with the lowest average index score, seem to be an exception to Gordon and Heath’s claim. However, this is reminiscent of Stuart-Smith’s working-class teenage girls (Figure 2.4), who also patterned with the males. Gordon and Heath call this patterning of the young, working-class / uneducated girls with the men the “tomboy pattern,” and following Eckert (1989:265) explain that such sociocultural divisions are sharper among the speech of the girls because they assert their category identities through language more than the boys. Of course, the social forces working behind the tomboy pattern observed by Eckert and others are most likely quite complex, and beyond the scope of our discussion. But this is another example of Gender Style.
Hypothesis 2: women use masculine variation that is above the level of gender awareness (a tomboy speech style) more often than men use feminine variation that is above the level of gender awareness.

To briefly summarize Gordon and Heath’s claims, women are attracted to the high front vowel while men are attracted to the back vowels because the high front vowel resonates at what is perceived to be a high pitch while the back vowels resonate at what is perceived to be a low pitch. High pitch, and hence the high front vowel, is associated with smallness and femininity while low pitch, and hence the back vowels, are associated with largeness and masculinity. Although the link between the variation and gender is indirect (i.e., there are no anatomical reasons for women to produce the high front vowel more than men), Gordon and Heath’s claims are compatible with the overall framework of engendered phonetic-level variation presented here. Labov concedes that there may be some explanatory value in Gordon and Heath’s proposal, but dismisses it as a theory of language change for two reasons. The first reason is that this explanation does not account for consonantal changes led by women, of which Labov refers specifically to the devoicing of /ʒ/ in Buenos Aires. The second reason is that there are exceptions to Gordon and Heath’s pattern of sound symbolism, of which Labov mentions “the female predominance in the mergers of vowels before /l/ [in the Southern Cities Vowel Shift], which proceed by laxing and centralization of the tense member of the opposition” (Labov 2001a:291). These two examples of putative counterevidence to Gordon and Heath’s theory of sound symbolism should be familiar; they are the same two examples Labov used when arguing against his generalization that women lead vowel changes that result in the increased dispersion of the vowel system. I have just discussed both of these examples in detail, and we saw that unlike the theory of sound symbolism, the phonetic-level engendered variation framework was able to account for the sex differentiation patterns seen there.

By making the distinction between engendered and nongendered phonetic-level variation as I am calling for in this dissertation, we are able to resolve the Gender Paradox, at least at the phonetic-level. The key to the paradoxical linguistic behaviour of women lies in the fact that we are talking about two levels of social awareness, and it is crucial that they not be confused. When phonetic-level changes take place below the
level of awareness, women show a different pattern because prestige is not a factor in determining their linguistic behaviour. Once we remove prestige from the equation, then if applicable, gender social significance takes over. This results in, for example, women leading phonetic-level changes that entail the preservation of distinctions (a feminine variant), and men leading changes that entail the reduction of distinctions (a masculine variant). The result is that on one hand women appear to be conservative when a change involves variants with prestige associated with them, in that they tend to go with the prestigious variant. But on the other hand, women appear to be innovative when the variants are not associated with prestige, since women tend to adopt the new variant. But if we think of this as the preservation of an old distinction instead of the adoption of a new variant, then we see that in the end, women are more conservative with respect to both changes above and changes below the level of social awareness.

4.5 Phonetic-level engendered variation and sociolinguistic theory

In Chapters Two and Three, I argued for the existence of engendered phonetic-level variation. The objective of this chapter was to illustrate the significance this type of variation has on sociolinguistic theory. Once we acknowledge that at the phonetic level, behavioural, anatomical and physiological differences between men and women play a role in determining how we speak, then we see that these differences also play an important role in phonetic-level language variation and change. If the variation is below the level of social awareness, then the social significance Labov calls “prestige” is not a factor in determining the linguistic behaviour of men and women. However, another type of social significance, one which I have called gender social significance is a factor. For example, working-class speakers attach social significance to sounding masculine, and therefore we predict that we will see phonetic-level engendered variation that correlates with social class, even if the variants are void of prestige. For this reason, a prestige-based model of style shifting is not an appropriate model with which to examine phonetic-level engendered variation.

The significance of the engendered versus nongendered distinction at the phonetic level is particularly relevant when it comes to theories of language change. In the last section of this chapter, I introduced three of Labov’s principles on the role of sex in language variation and change. I then argued that the third principle — women lead
linguistic changes that occur at below the level of social consciousness — does not hold at the phonetic level with respect to engendered variation. I reviewed several well-known counterexamples to this principle, and resolved these counterexamples within a phonetic-level engendered variation framework. Lastly, I attempted to explain Labov’s Gender Paradox by claiming that from a behavioural phonetic-level engendered variation perspective, the linguistic behaviour of women in changes from below the level of social awareness is actually conservative, since they are preserving phonetic-level distinctions. Thus we are able to label their linguistic behaviour as “conservative” during both changes from above and changes from below. These insights into the way language changes at the phonetic level motivate making the engendered–nongendered distinction.

I have tried to capture the essence of the role these sex differences play in phonetic-level variation in several hypotheses. I doubt that the few hypotheses put forward in this chapter are an exhaustive list of the role engendered variation plays in language variation and change, but they should be sufficient to make researchers realize the significance of the phonetic-level engendered variation framework.
Engendered Variation in the Voices of Radio DJs

Speech reveals the man, but voice reveals the inner man.
Murphy 1964:24

5.1 Chapter objective

The objective of this chapter is to test empirically the most fundamental claim made in the previous chapters, viz., that the two types of sex-based variation introduced in Chapters Two and Three form a frame for socially significant gender-based variation. With regard to anatomically-driven sex-based variation, this claim was articulated as a characteristic of engendered variation. The list of characteristics of engendered variation presented on page 59 is repeated here:

- when enough speakers are looked at, sex-based linguistic patterns emerge
- these patterns tend to be consistent across different speech communities
- the way each of the sexes patterns is grounded in the sex differences in the anatomy of the vocal tract
- the extent of the sex differentiation seen in the engendered variation varies from one speech group / community to the next, but the direction the sexes pattern tends to remain the same
- however, there are individuals or groups who for social reasons pattern more like the opposite sex

The last characteristic is the relevant one for our purposes: there are individuals or groups who for social reasons, i.e., their social gender, pattern more like the opposite sex. In other words, the sex-based variation sets up a frame for the gendered-based variation so
that, for example, if men typically produce a certain variant more often than women, then individuals who highly value masculinity will typically produce the male variant more often than those individuals who do not highly value masculinity. To give a concrete example, if in general men voice /h/ more often than women, then men who highly value masculinity will voice /h/ more often than men who do not highly value masculinity.

I am not the first to suggest that sex-based anatomical variation forms a frame for gender-based variation. As I discussed in Chapter Two, Stuart-Smith (Stuart-Smith et al. 2003), in her examination of the variation seen in the production of sibilants in Glasgow, provides empirical evidence for a similar framing of engendered variation by sex based anatomical differences in the vocal tract.

In a similar fashion, I intend for behaviourally-determined sex-based patterns to also provide a frame for behaviourally-determined engendered variation. This was formulated as the *Gender Hypothesis 1*:

**Gender Hypothesis 2 (page 86)**

Phonetic distinctiveness indexes social gender.

The extension of the gender frame to encompass both behavioural differences and anatomical differences is new, and is the main contribution of this chapter. Similar to the anatomical frame, I also intend for the relationship between behaviourally-determined sex-based variation and behaviourally-determined gender-based variation to hold across languages, and for all types of sex-based behavioural differences. Again, it is difficult to test the universality of the cross-linguistic consistency of the behaviourally-determined variation. Instead I will present one example, contextual and inherent vowel length contrasts in English, in detail.

We will also be testing the first of the social class hypotheses, which is repeated here:

**Social Class Hypothesis 1 (page 92)**

Working class speakers use more masculine variants than middle class speakers, while middle class speakers use more feminine variants than working class speakers.
5.2 Recording gender on tape: The radio DJ mini corpus

5.2.1 Why use radio DJs?

It is not a simple matter to capture a wide range of social gender on tape. A number of experimenters have used mostly, or in some cases exclusively, university students in their attempts (for example, Biemans 2000). This approach is problematic because gender is closely related to social class (as discussed in §4.2), and therefore the range of gender is limited by selecting what will be mostly speakers with a middle-class background.

By studying the voices of radio disc jockeys, I eliminate the problem of a single class by looking at a range of genres. There are very strong social class associations with different genres of music. The stereotypes are clearly seen in popular associations of classical music with gentility and haute couture, and heavy metal with toughness and earthiness. By using the voices of radio DJs from a number of different genres, we tap into these stereotypical associations.

This approach is not without its challenges. Unlike the typical sociolinguistic interview, we do not have direct access to the DJs themselves, and therefore cannot ask questions that reveal the social background of the DJ. Some information about the DJs was retrieved from the internet, and is provided in §5.2.3. More important than their social characteristics are their on-air personalities. Anthropological linguistics studies (Bell 1982, Coupland 1985, Coupland 2001, but cf. Cutillas-Espinosa and Hernández-Campoy 2006 for limitations to their claims) of variation in the voices and speech of radio DJs and newscasters shows that they “alter their style of speech depending upon who they think is listening” (Bell 1982:150). In other words, the DJs perform stereotypical roles that conform to the expectations of the audience. Without the audience, we would not get the performance. Using the speech of DJs recorded while they are on the air allows us to study a wide range of gender-based identities.

I base the perceived social characteristics of the DJs on evaluations of their voices by naïve listeners. (The perception experiment is discussed in detail in §5.3.) This approach allows me to capture a broad range of gender variation, as necessary to test the hypotheses.
5.2.2 Selecting and recording the DJs

All of the DJs work for the same company, XM Satellite Radio. This company broadcasts from their studio in Washington, D.C., to three satellites in geosynchronous orbit above North America. The satellites repeat the signal and project it back to earth, allowing for reception of the signal in most of North America. XM Satellite Radio broadcasts over 170 stations that cover all genres of music, as well as news, sports, weather, and talk shows.

The stations do not play commercials. Instead, they make a profit by charging listeners a monthly subscription fee. Instead of commercials, the music or talk is often separated by in-house promos. These promos are of two types. They either inform the listener of special events, such as the guest appearance of a musician on another station of a similar genre, or they remind the listener of the name and channel number of the station they are currently listening to.

With so many stations, we have a good variety of voices to choose from. I reduced the selection by considering only white males on the music stations. This unfortunately eliminated some of the music genres such as rap and soul. However, in order to argue that the variation heard in the voices of the DJ is a reflection their social gender, it is necessary to eliminate as many other sources of variation as possible, of which one is ethnicity. DJs who consistently played music in the background while they talked were also eliminated. This was the case for all of the dance music and techno-style DJs.

From the remaining DJs, I selected eight. These eight DJs and their music genres are summarized in Table 5.1. I provide more biographical information about each DJ in the following section. Altogether six genres are represented: classical, country and western, popular, hard rock, punk, and heavy metal. Two of the genres, classical and country and western, are represented by two DJs. Many of the stations have only one radio personality. When that individual is not in the studio, then the station airs a pre-programmed selection, or reruns a show that was aired earlier in the week.

I recorded the speech samples of the DJs from a Pioneer XM2GO satellite radio. The signal is digitized at the studio. Although the radio is a digital receiver, the only output is an analog headphone jack. I suspect the company avoids a digital output on the radio in order to protect the itself from pirating. The analog signal was re-digitized using
a IBM compatible desktop PC running PRAAT version 4.2.2 with a sampling frequency of 22,050 Hz. I recorded approximately 60 minutes of speech for each of the DJs. On-air episodes by DJs lasted from as short as five seconds to as long as about five minutes. The DJs talked on an average of eight minutes per hour of show time. The recording process was spread out over several months during the summer and fall of 2006. My impression is that all of the DJs are very consistent from day to day in their speech style and other speech patterns.

5.2.3 Biographical information for the eight DJs

Following is some biographical and career information for a subset of the DJs. Some of these DJs are well-known, and therefore there is information about them available to the public. Others are obscure, and it was impossible to find any information on them.

Martin Goldsmith was born in St. Louis, where his mother spent 21 years as a member of the St. Louis Symphony Orchestra. His father was also a professional musician. Goldsmith graduated from Johns Hopkins University in Baltimore. While a student, he sang in the Baltimore Opera Company. He has also acted in many roles in Washington-area theatres. He began his radio career at a commercial classical station in Cleveland. Between 1989 and 1999, he served as the host of Performance Today, the National Public Radio's daily classical music program. During that time, the show won several awards. He is now director of the classical music program at XM Satellite Radio.

<table>
<thead>
<tr>
<th>DJ Personality</th>
<th>station name</th>
<th>genre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martin Goldsmith</td>
<td>XM Classics</td>
<td>classical</td>
</tr>
<tr>
<td>Paul Bachman</td>
<td>XM Classics</td>
<td>classical</td>
</tr>
<tr>
<td>Country Dan</td>
<td>America</td>
<td>country &amp; western</td>
</tr>
<tr>
<td>J.W.</td>
<td>America</td>
<td>country &amp; western</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>U-POP</td>
<td>international pop</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>The Boneyard</td>
<td>80's hard rock</td>
</tr>
<tr>
<td>Lars Fredericksen</td>
<td>Fungus</td>
<td>punk</td>
</tr>
<tr>
<td>Coolguy</td>
<td>XM Liquid Metal</td>
<td>heavy metal</td>
</tr>
</tbody>
</table>

Table 5.1 The eight DJs, their affiliated station, and their music genres
He has also written two books, one about his parents and their experiences playing in an all-Jewish orchestra in Nazi Germany, and one about The Beatles.

Country Dan Dixon lived and worked in Michigan as a country and western music radio disc jockey before joining XM Satellite Radio. He started his career over 25 years ago.

Ted Kelly has been in the communications industry all his working life. He has held positions such as executive with CBS. He started his broadcasting career on Spanish radio, where discovered an affinity for pop music produced outside of North America. Kelly entered the market of digital media by creating online streaming media broadcasts for sports and news for CBS. He is also the managing director of Intervox Communications, a streaming media and broadcast consulting firm in Washington, D.C.

Chris Jericho, whose real name is Christopher Keith Irvine, was born in Manhasset, New York, but grew up in Winnipeg, Canada. He is currently living in Los Angeles, California, and Tampa, Florida. He has worked as an actor, a radio host, and a musician. He recently starred in a made-for television science fiction movie. Chris is best known for his role as Chris Jericho, a professional wrestler for World Wrestling Entertainment. In 2001 he became the “Undisputed Champion” of WWE. In the summer of 2005, Jericho retired from wrestling to pursue his career in the music industry. He is currently the lead singer of the rock band Fozzy. They have released three albums, of which the first two have sold more than 100,000 copies. Jericho directed the bands music videos, all of which have been played on MTV and Much Music. Jericho also owns the record label, Ash Records, that produces the band’s albums. He is a guest DJ at XM Satellite Radio, hosting a three hour show titled “The Rock of Jericho” once a week. The show focuses on lesser-known hard rock bands from North America and Europe. On the show, Chris displays a vast knowledge of the music techniques and history of the bands that he plays on the radio show.

Lars Frederiksen (born Lars Everett Dapello in Campbell, California in 1971) is a guitarist and vocalist for the punk rock band Rancid, as well as for Lars Frederiksen and the Bastards. Both of these are ska-punk bands. He has worked as a producer for bands such as the Dropkick Murphys. He joined Rancid in 1993, and to date, he is still a member of the band. Frederiksen is half-Danish (his mother, Millia, was a Danish
immigrant) and was raised in Campbell, California, a suburb of San Jose. He attended Westmont High School, but he did not graduate.

5.2.4 Radio DJs and the non-standard–standard speech continuum

The DJs show socially significant variation along two related continua, the standard–non-standard continuum, and what I will label as the engendered–nongendered continuum.

I will first illustrate of the standard–non-standard continuum beginning with a twenty-second excerpt of speech from Country Dan, one of the country and western DJs.

This is XM ten America. That one requested by Wanderin’ Coyote, who sent me a beautiful picture of where he’s at, from his cell phone of course. And he’s out there in Colorado headed to Nebraska. The only question I have for you, Wanderin’ Coyote, is that birds out there, or is that bugs on your windshield? I couldn’t tell.

Although there is more than one example of non-standard English here, I will discuss only one variable, the pronunciation of the participle suffix -ing. As is well known, in careful styles of speech the participle suffix is pronounced with a final velar nasal consonant. The velar consonant is representative of middle-class speech, and as such is considered the standard form. The casual variant is the alveolar nasal. This variant is more common both in casual speech and working-class speech, and is considered the non-standard variant. The sample of speech from Country Dan shown here contains two occurrences of the participle suffix, both the non-standard (indicated orthographically by the apostrophe).

At the other end of the standard–non-standard continuum we have the classical DJ Paul Bachman. Following is a fourteen-second excerpt of his speech.

A recording from 2004, one that is coupled with the Elgar violin concerto, featuring violinist Hillarie Han, joined by the London Symphony Orchestra. Sir Colin Davis conducting.

This excerpt contains three occurrences of the suffix –ing, two as a participle and one as a gerund, all of which are the standard form. A comparison of the choice of variant between these two DJs suggests that there is a large amount of variation in their speech. But how representative of the speech of the DJs are these two samples? Not only are they very short, but we also have not controlled for grammatical category, something that has a very large influence on the choice of variant (Houston 1991). If it turns out that these
examples exaggerate the range of variation seen in the speech of the DJs along the standard–non-standard continuum, then it is doubtful that we captured a wide range of variation along the engendered–nongendered continuum, as the two are closely related.

In order to validate the broad range of variation in the speech of the DJs, the first 50 tokens of the participle suffix were extracted for each DJ. Grammatical category was controlled for by extracting only present participles (and not nouns, adjectives or gerunds). This context favours the non-standard variant. Figure 5.1 shows the percentage of the tokens that are non-standard for each of the DJs. The percentages range from zero for Martin Goldsmith to 100 percent for Coolguy. This large range of diversity in the usage of the standard and non-standard variants establishes that there is a broad range of variation along the standard–non-standard continuum in the voices of the DJs. Furthermore, the fact that one DJ uses exclusively the non-standard variant while another DJ uses exclusively the standard variant suggests that the DJs’ speech styles contain a performance component, and it is extreme or stereotypical. With this evidence, we are now ready to consider the engendered–nongendered continuum.

5.2.5 The DJs and different facets of masculinity

Besides the standard–non-standard continuum, the DJs also vary along an engendered–nongendered continuum. Speech at the engendered end of the continuum contains a variety of linguistic and paralinguistic indexations of social gender, in this case masculinity. The nongendered end of the continuum is void of any indexations of social gender (although in practice speech void of gender indexations is most likely unattainable by an adult).

As discussed in §4.2.1, masculinity is a multi-faceted personality trait that manifests itself in a number of ways. According to Connell (2005), three of the primary components of masculinity are machismo (physical size and strength, martial ability, etc.); technical ability and knowledge, such as the ability to repair motor vehicles; and sexual vigour.

The different DJs display different components of masculinity, and they do so in a number of different ways. Phonetic-level engendered variation is one of those ways. Other possibilities, all of which will be illustrated shortly include displays of toughness,
ENGENDERED VARIATION IN THE VOICES OF RADIO DJS

Consider first the following example of machismo. Here, the DJ Coolguy is reflecting on how his attempt to make fun of the talk show called the Opie and Anthony Show, which caters to an adult male audience (with frequent sexually-explicit references), initially backfired:

I started saying some things on the Boneyard, when I was on the Boneyard, you know, and they [i.e., Opie and Anthony] responded, and well they, uh, they beat me up. They beat me up pretty good. But I could take it, and I deserved it anyway. Since then things have changed, I’ve gotten on the show’s good side, especially Jim Norton, who’s part of the Opie and Anthony Show, and, uh, people are asking me when I’m gonna be on their show.

Opie and Anthony responded to Coolguy’s comments by verbally “beating” him up. This double-edged image of mutual aggressiveness — both the ability to give it as well as the ability to take it — ended with a renewed mutual admiration, and an invitation from Opie and Anthony to Coolguy to appear on their talk show. Coolguy is representing himself as a macho individual — someone capable of giving and receiving punishment. Although the punishment is not literally physical, the connotations are there; if it came to physical blows, Coolguy would not be one to back down.

The second aspect of masculinity that I mentioned is technical ability. In the case

References to sex, and the use of profanity.

Consider first the following example of machismo. Here, the DJ Coolguy is reflecting on how his attempt to make fun of the talk show called the Opie and Anthony Show, which caters to an adult male audience (with frequent sexually-explicit references), initially backfired:

I started saying some things on the Boneyard, when I was on the Boneyard, you know, and they [i.e., Opie and Anthony] responded, and well they, uh, they beat me up. They beat me up pretty good. But I could take it, and I deserved it anyway. Since then things have changed, I’ve gotten on the show’s good side, especially Jim Norton, who’s part of the Opie and Anthony Show, and, uh, people are asking me when I’m gonna be on their show.

Opie and Anthony responded to Coolguy’s comments by verbally “beating” him up. This double-edged image of mutual aggressiveness — both the ability to give it as well as the ability to take it — ended with a renewed mutual admiration, and an invitation from Opie and Anthony to Coolguy to appear on their talk show. Coolguy is representing himself as a macho individual — someone capable of giving and receiving punishment. Although the punishment is not literally physical, the connotations are there; if it came to physical blows, Coolguy would not be one to back down.

The second aspect of masculinity that I mentioned is technical ability. In the case

References to sex, and the use of profanity.
ENGENDERED VARIATION IN THE VOICES OF RADIO DJs

of radio DJs introducing music, technical ability equates to detailed knowledge about the bands, their histories, and their musical techniques. The following excerpt from Chris Jericho, is illustrative of the type of technical knowledge displayed by the DJs. Here Chris compares the different playing styles of two of the drummers for the heavy metal band, KISS:

I wanted to play some heavy Eric Carr because I saw the Crazy Nights video the other day on VH1’s Metal Mania, which is a great show. You should watch it if you are a fan of the Rock, and I just remember thinkin’, man, fuckin’ Eric Carr’s drumming is so heavy, he is so hard-hitting, he is the exact antithesis, the exact opposite of what Peter Criss was doing, when he was in KISS, and I can really see how they just would totally dug the fact that they had this hard-hitting, pounding rock and roll drummer after having kind of the swing style of Peter Criss, who could still play, I mean you heard his drum solo from 100,000 Years from KISS Alive a couple weeks ago. There is no doubt about that, but there’s just two different styles, and for my money, I dig Eric Carr’s style. It’s just, it’s more of a heavy metal type sound.

Jericho is showing his technical knowledge of the heavy metal music industry by being able to discuss the playing style of the former drummer for the band KISS. He contrasts and evaluates the style of the former drummer with the style of the current drummer. His knowledge in this short excerpt is representative of his show in general, and my overall impression is that he commands a level of knowledge that far exceeds that of the average listener.

The third facet of masculinity is sexual vigour. Men, in general, talk about sex and use sexually-explicit profanity more often than women. Thus, discussions about attracting women and the use of explicit language is another way to index masculinity. This facet of masculinity is common, and often uninhibited. The following excerpt is the DJ Coolguy taking a request from a listener:

Coolguy: It’s 42, Liquid Metal, love it man. [indistinguishable speech] that’s [indistinguishable speech], The Killer, also Metallica with Fade to Black. One of the best right there, man. Love that tune. Eight-six-six-four-six-six-XM LM. How are ya?

Caller: Huh?

Coolguy: Never mind. How you doin’? You li- You diggin’ the channel?

Caller: Yeah, hell yeah, man.

Coolguy: How old are ya?
Caller: Twenty-two.

Coolguy: Fuckin’ A. I used to be twenty-two. Remember when you were nine?

Caller: Yeah, hell yeah, man.

Coolguy: So do I. That was a good time.

Caller: We fucked it up, man. Me and my cousin, we tore it up.

Coolguy: What you and your cousin, you used to fuck cattle too, like I did at that age?

The caller does not respond.

Coolguy: Never mind. What can I do for ya?

Coolguy then proceeds to take the song request from the caller. He also asks for the caller’s first name and the name of the city he is calling from. Coolguy lists a number of other songs that will also be played in the coming set, and asks the caller if this is all right. Coolguy then concludes the conversation by returning back to the topic of sex with an animal:

Coolguy: And maybe later after the show we could find a blind heifer somewhere, chase it around the barn yard, maybe hold it down and we could-

Caller hangs up.

Coolguy: Hey? What the fuck?

My impression of this conversation is that the caller is attempting to accommodate to the speech of Coolguy by using profanity. However, Coolguy feels the need to assert his own masculinity, and he ups the ante by introducing the topic of sex with an animal, to which the caller is unable to respond.

The previous example is one of the most extreme sexual references among my recordings. Following is a milder example. This conversation takes place between Ted Kelly and John, Ted’s co-host. They are commenting on the dress of some of spectators at the World Cup soccer tournaments:

Ted: It’s Ted Kelly, John Dowan, coast to coast in America. We’re satellite radio’s longest running breakfast show. It’s Ted Kelly’s world party, stateside in America on XM, across this planet on world space satellite radio. So you know, everybody is bitchin’ and complainin’ about these soccer hooligans causin’ all the problems in Germany, all the issues, right?
John: Yeah, sure.
Ted: With so many hot girls dressed in bikinis, with their nations’ colours painted on them in provocative ways, who even has time to worry about soccer hooligans?
John: Who has time to follow the games?
Ted: You got a point there.

One of the selling points of satellite radio is its freedom from censorship. As a result, it is not unusual to hear such explicit sexual references on radio. It is clear that the company institutionalizes the use of such language. Every station has several promos that are played between songs to remind the listener what station they are listening to, and to promote the general ambiance of the station. The following promo is from XM 41, the hard rock station called The Boneyard:

[In a female voice] Then there’s Bone. XM four-one, The Boneyard. She still had her high heels on. He wondered if she was going to keep them on in bed. Was she in to kinky sex? Bone me!

The Boneyard plays 1980s hard rock music, such as Iron Maiden, Dio, and Ozzy Osborne. These bands are for the most part long gone; all that remains are skeletons, that is, bones, that get dug up and played on the air. The last words of the promo are a play on the word “bone,” which as a verb refers to the act of having sex and the objectification of women. This promo is similar to many of the promos played on this station in that they almost always make a sexual reference in some manner or form. For stations like The Boneyard, the general atmosphere at the company encourages the DJs’ use of sexual references on the air.9

Sex-related profanity is a related facet of masculinity. Its use, particularly terms that objectify women (such as “bone me”) serve the important function of preserving male / female differences in status and power (Kutner and Brogan 1974). Males who use these terms frequently are showing their interlocutors that they are the dominant role in relationships with the opposite sex. The maintenance of this power imbalance is obviously flattering to males, and might be expected to attract masculine audiences who share these values. The following excerpt illustrates the use of sex-related profanity by the DJ Lars Fredericksen. He and his co-host, Gordon Cambodie, are introducing the

9 Of the 68 music channels broadcasted by SM Satellite Radio, seven of them have explicit language warnings. Concerned parents can block the reception of these channels.
songs that they just played on the air:

Lars: This is Fredericksen. Gonna tell ya what ya heard. Kicked it all off with a band called Cranked Up. That was a track called “Deal With It.” Unfuckin’ real, these guys.

Gordon: That was a really fuckin’ good track, man.

Lars: Really fuckin’ really good you guys. Like unfuckin’ believable. Ummm… kinda had that like… ummm… I wanna say… ah, shit… kinda- I wanna- not U.S. Bombs, but kinda Antiheroes.

Gordon: I could see that. I could see that.

Lars: But for me, and I don’t know, fucking really really good.

Gordon: Great- just a great song, all around.

Lars: Yeah, great song. Street Brats after that.

Gordon: Street Brats are fucking solid, Austin, Texas. Fuckin’ just good motherfuckers, know what I mean, these guys have been in so many bands, and it shows. So fuckin’ good man.

Both Fredericksen and his co-host make extensive use of profanity, particularly the word “fuck” in its various forms, and in this way are indexing this facet of their masculinity.

In order to represent the opposite pole of language that the DJs use, I will conclude with an excerpt from the classical DJ, Martin Goldsmith:

I am Martin Goldsmith, with the music for you here on channel 110, XM classics, coming to you from XM satellite radio. Let’s remain in America with Symphony Number One, by Charles Ives, music that the soon-to-become-iconoclastic American composer wrote while he was still in college, at Yale University, in New Haven, Connecticut. This first symphony is very much a product of the late Romantic era. It’s played for us by the Detroit Symphony Orchestra with Neeme Järvi conducting. The Symphony Number One, by Charles Ives, here on XM Classics.

This quotation contains details about the composer and the composition, but it is completely void of displays of machismo, references to sex, profanity, or non-standard speech.

The different excerpts from the DJs presented in this section showcase the multifaceted nature of masculinity, and illustrate ways in which the DJs tap into and make use the aspects to masculinity, be it machismo, technical knowledge, sexual references, or profanity. Thus the speech of the DJs provides a testing ground for our hypothesis on the role of social gender in language variation.
5.3 The gender perception experiment

In order to assess the gender constructions of the speakers with a degree of objectivity, I rely on the evaluations from naïve listeners. This allows for an objective, quantitative measure of the perceived gender differences of the DJs (Chambers 2000). After quantifying their gender in this manner, we can test the significance of correlations with linguistic variation in their speech.

5.3.1 The experiment

Selection of speech samples

One minute samples were selected for each DJ. Transcriptions of the speech samples are provided in Appendix B. Wherever possible, I selected continuous monologues to avoid discontinuities in the flow of speech and to avoid confusion by the listener about the voice they are rating. However, not all of the DJs produced one minute monologues, and one DJ, Coolguy, rarely produced a monologue over 10 seconds. For these DJs, multiple monologues were used. For Coolguy, a combination of monologue and dialogue was used. The monologues were presented first to allow the listener to become familiar with his voice. Otherwise, speech samples were selected for their low background noise levels and my impression of the sample’s representativeness of the DJ’s speech in general. No attempt was made to conceal the genre of the music that the DJ was playing.

Selection of participants

Participants were recruited from the undergraduate population at the University of Toronto. I only recruited students who were born and grew up in Canada or the United States, were native speakers of English, did not consider themselves members of an ethnic minority, and did not have any hearing problems. Eight women and five men participated in the experiment.

The perception scales

The participants were instructed to rate each voice on fourteen seven-point Likert scales. The attributes rated, a brief description, and the end-point labels were provided to the participants before they began the experiment (Table 5.2). The participants were encouraged to ask questions about the scales if they found the description to be unclear. The names of the scales and the end point labels were reproduced on each of the
questionnaire sheets to remind the listener of what they were measuring. The left label corresponds to a 1 on the scale, while the right label corresponds to a 7.

Of the fourteen scales listed in Table 5.2, *Sense of Humour* is a distractor, and is not intended to explain anything. Two others, *Clarity of Speech* and *Voice Quality*, are

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Scale / List Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>What do think the age bracket of the DJ is?</td>
<td>teens 20s 30s 40s 50s 60s 70s</td>
</tr>
<tr>
<td>2. Regional Accent</td>
<td>How similar is the DJ’s accent to that of people from Toronto?</td>
<td>definitely from around here</td>
</tr>
<tr>
<td>3. Sense of Humour</td>
<td>Does the DJ sound like he has a sense of humour and likes jokes?</td>
<td>funny; prankster</td>
</tr>
<tr>
<td>4. Technical Ability</td>
<td>Does the DJ sound like he can fix cars, computers, etc.?</td>
<td>can repair cars and computers</td>
</tr>
<tr>
<td>5. Education</td>
<td>How much education do you think the DJ has completed?</td>
<td>high school dropout</td>
</tr>
<tr>
<td>6. Knowledge</td>
<td>How knowledgeable of music and its history does the DJ sound?</td>
<td>has encyclopedic knowledge of music</td>
</tr>
<tr>
<td>7. Friendliness</td>
<td>How friendly does the DJ sound?</td>
<td>very friendly; compassionate</td>
</tr>
<tr>
<td>8. Macho</td>
<td>Does the DJ sound like a person who would win a bar fight?</td>
<td>tough; macho</td>
</tr>
<tr>
<td>9. Physical Size</td>
<td>Does the DJ sound like he a physically large person?</td>
<td>hefty; bull-like</td>
</tr>
<tr>
<td>10. Clarity of Speech</td>
<td>Is the speech of the DJ clear and easy to understand?</td>
<td>very easy to understand</td>
</tr>
<tr>
<td>11. Composure</td>
<td>Does the DJ sound relaxed?</td>
<td>laid-back relaxed</td>
</tr>
<tr>
<td>12. Voice Quality</td>
<td>Does the voice of the DJ sound smooth or rough?</td>
<td>smooth; clear</td>
</tr>
<tr>
<td>13. Material Wealth</td>
<td>Do you think that the DJ is rich?</td>
<td>a billionaire</td>
</tr>
<tr>
<td>14. Self-confidence</td>
<td>Does the DJ sound confident in himself?</td>
<td>overly confident, arrogant</td>
</tr>
</tbody>
</table>

Table 5.2 The fourteen Likert scales used in the gender perception experiment. For each scale, the name, a brief description, and the labels at each end point are also provided.
not social characteristics, and were not intended for use in the correlations. They were included to see if the listeners were able to perceive behavioural linguistic differences in the voices of the DJs. The remaining eleven scales are designed to elicit the perceived social and personality characteristics of the DJs.

**Presentation of the speech samples**

The intensity of the speech samples was adjusted so that they were within 1 dB of each other using the Amplify function of Audacity ver. 1.2.4. The modified samples varied from 56.99 dB to 57.90 dB. The samples were presented in sequence over Panasonic RP-SP 30 speakers at a comfortable listening volume. The participant was instructed to not wait until the sample ended before they began rating the voice. After the one minute sample finished, the participant was given as much time as they required completing the ratings for that DJ before proceeding to the next voice sample. They were instructed to complete all the ratings, and to assign a value of four if they were completely unsure of how to rate the voice on a particular scale. Altogether, there were five occurrences of missing ratings, three for *Age*, one for *Regional Accent*, and one for *Sense of Humour*.

**5.3.2 The initial results**

The means and standard deviations averaged across the responses for the thirteen participants are in Table 5.3. A comparison of the standard deviations by DJ reveals that the participants showed the most agreement on their judgments of Ted Kelly (average standard deviation = 0.90), and the least agreement on their judgments of Lars Fredericksen (average standard deviation = 1.37). A comparison of the standard deviations by Likert scale reveals that the participants showed the most agreement for their perception of age (average standard deviation = 0.55) and material wealth (average standard deviation = 0.77), and the least agreement for regional accent (average standard deviation = 1.39).

I ran a repeated measures one-way ANOVA to assess if the listeners were able to differentiate between the DJs’ voices on the eleven scales I intended to use as social variables. For each analysis, the assumption of sphericity was tested for with a Mauchly test (Max and Onghena 1999), and if violated the degrees of freedom were adjusted with a Greenhouse-Geisser correction. Table 5.4 reports the $F$ statistic, p level, degrees of
freedom, and which scales violated the assumption of sphericity. The perceived differences between the DJs are significant for all of the scales except Technical Ability. An examination of Technical Ability in Table 5.3 shows that the means hover around four, the value the subjects gave for complete uncertainty. This suggests that the participants were not able to differentiate the technical ability of the DJs based on the one minute speech samples because they lacked a perception of technical ability altogether (as opposed to, for example, all of the DJs being rated as having a high level or lower level of technical ability). Technical Ability was therefore removed from the analysis. This left ten scales to be used as the social characteristics of the DJs.
5.3.3 Aggregating the scales with Principal Components Analysis

The Likert scales are designed to capture different aspects of the same general social characteristics of the DJs. We thus expect there to be correlations between the scales. For example, **Machismo** and **Physical Size**, both aspects of masculinity, will most likely correlate with each other. The same is true for **Material Wealth** and **Education** (at least for Americans), two aspects of the perception of social class. Using predictors that covary with each other in this manner to explain the variation seen in the dependent variable runs the risk of multicollinearity. The redundancy in the independent variables results in overfitting the regression analysis model, and distorts the results. To avoid this overfitting of the model, we need to aggregate the scales that correlate with each other in meaningful ways.

Not all of the scales are expected to group with others. Any correlations with **Age** or **Regional Accent** would be coincidental. **Age** and **Material Wealth** may correlate with each other (and they do, $r(99) = .478$, $p < .000$) because in general older people have

<table>
<thead>
<tr>
<th></th>
<th>Correction</th>
<th>d.f. between groups</th>
<th>d.f. within groups</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>sphericity assumed</td>
<td>7</td>
<td>63</td>
<td>22.48</td>
<td>.000</td>
</tr>
<tr>
<td>Regional Accent</td>
<td>sphericity assumed</td>
<td>7</td>
<td>77</td>
<td>14.01</td>
<td>.000</td>
</tr>
<tr>
<td>Technical Ability</td>
<td>sphericity assumed</td>
<td>7</td>
<td>84</td>
<td>1.14</td>
<td>.316</td>
</tr>
<tr>
<td>Education</td>
<td>sphericity assumed</td>
<td>7</td>
<td>84</td>
<td>24.74</td>
<td>.000</td>
</tr>
<tr>
<td>Knowledge</td>
<td>sphericity assumed</td>
<td>7</td>
<td>84</td>
<td>4.77</td>
<td>.000</td>
</tr>
<tr>
<td>Friendliness</td>
<td>sphericity assumed</td>
<td>7</td>
<td>84</td>
<td>3.72</td>
<td>.002</td>
</tr>
<tr>
<td>Machismo</td>
<td>sphericity assumed</td>
<td>7</td>
<td>84</td>
<td>13.07</td>
<td>.000</td>
</tr>
<tr>
<td>Physical Size</td>
<td>Greenhouse-Geisser</td>
<td>3.45</td>
<td>41.45</td>
<td>8.65</td>
<td>.000</td>
</tr>
<tr>
<td>Composure</td>
<td>Greenhouse-Geisser</td>
<td>3.98</td>
<td>47.73</td>
<td>4.31</td>
<td>.005</td>
</tr>
<tr>
<td>Material Wealth</td>
<td>Greenhouse-Geisser</td>
<td>3.26</td>
<td>39.08</td>
<td>13.99</td>
<td>.000</td>
</tr>
<tr>
<td>Self-confidence</td>
<td>sphericity assumed</td>
<td>7</td>
<td>84</td>
<td>6.44</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 5.4 Repeated measures one-way ANOVA comparing the Likert scale responses by the 13 participants across the eight DJs
accumulated more wealth than younger people. Furthermore, in this particular case it is coincidental that the Classical DJs, who use a more standard speech style, and therefore were predictably perceived as more wealthy, are also older than the other DJs.

A Principal Components Analysis with varimax rotation was conducted to look for relationships among the following eight scales: Education, Knowledge, Friendliness, Machismo, Physical Size, Composure, Material Wealth, and Self-confidence. These eight scales reduced to three components accounting for 74.7 percent of the variation after rotation. Table 5.5 shows the loading weights for each of the scales, with weights less than .35 omitted for clarity. Only the scales with loading weights listed under a component are considered to be a contributor. The first component decreases as Education increases, and increases as Machismo, Physical Size, and Self-confidence increase. Of the four scales that contribute to this component, Education contributes the least. This component reflects the perceived masculinity of the DJs, the higher the score the greater the machismo facet of masculinity. The second component increases as Education, Knowledge, and Material Wealth increase. This component reflects the perceived social class of the speaker, the higher the score, the higher the social class. The third component increases as Friendliness and Composure increase. This component

<table>
<thead>
<tr>
<th></th>
<th>Component One</th>
<th>Component Two</th>
<th>Component Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>-.51</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>Friendliness</td>
<td></td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Machismo</td>
<td>.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Size</td>
<td>.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composure</td>
<td></td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Material Wealth</td>
<td></td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Self-confidence</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.5 Loading weights for the three components underlying the perception data. The first component consists of the scales that reflect Social Gender, the second Social Class, and the third Personality.
reflects the perceived personality of the DJs.

Following the results of the Principal Components Analysis, the scales were aggregated to create three general scales called Gender, Social Class, and Personality. The general scales were created by summing the products of the contributing Likert scales with their corresponding loading weights in Table 5.5 (Woods et al. 1986:275). In the case of a negative contributor, the reverse of the scale (8 minus the value) was used. Following are the formulae for the three aggregate scales:

Gender = \(0.51*(8-\text{Eduation}) + 0.82*\text{Machismo} + 0.84*\text{Physical Size} + 0.77*\text{Self-Confidence}\)

Social Class = \(0.78*\text{Education} + 0.79*\text{Knowledge} + 0.87*\text{Material Wealth}\)

Personality = \(0.87*\text{Friendliness} + 0.80*\text{Composure}\)

Table 5.6 lists the aggregated values for the eight DJs.

We now have a complete set of perceived social characteristics for the eight DJs, consisting of Age, Regional Accent, Gender, Social Class, and Personality. These values will act as independent variables in our tests for relationships with the dependent linguistic variable, the amount of word-final sibilant devoicing. To this list, we will add independent linguistic variables, such as phonological environment, as appropriate.

5.4 Word-final sibilant devoicing in English

5.4.1 Devoicing as sex-based variation

In this section, I examine sibilant devoicing in the voices of the eight DJs, with the intention of demonstrating that anatomically-grounded sex-based differences form a frame for gender-based differences. This study looks exclusively the devoicing of /z/ in word-final position. I only focus on /z/ for two reasons. First, several studies of alveolar fricatives (see §2.3 for review) have found that they are capable of indexing social information such as social gender. Second, the alveolar fricative is by far the most common fricative to occur in word-final position, and thus it will be relative easy to find a sufficient number of tokens in natural speech.

The segments /s/ and /z/ in English contrast for phonological voicing. This phonological contrast is realized in a number of ways (following Slis and Cohen 1969):
the presence or absence of periodicity

the duration of the fricative

F0 patterns in the following vowel

the durations of the formant transitions into the following vowel

the duration of the preceding vowel

Of these, periodicity is perceptually the most important (Stevens et al. 1992). In general, the periodic vibration of the glottis does not occur throughout the entire duration of the fricative, but rather devoicing occurs as time progresses, so that the beginning of the fricative is voiced and the end is voiceless, that is, the vocal folds have ceased to vibrate. This results in a gradual shift from voiced to devoiced. The extent of devoicing depends on the following factors (Haggard 1978):

- syllable position
- place of articulation
- voicing characteristics of the adjacent segments

Fricatives in word-final position devoice more than those in word-initial position, and are

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Social Class</th>
<th>Personality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Dan</td>
<td>14.6</td>
<td>9.6</td>
<td>9.4</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>13.2</td>
<td>12.2</td>
<td>7.4</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>12.6</td>
<td>9.2</td>
<td>6.4</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>11.7</td>
<td>11.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>10.4</td>
<td>14.2</td>
<td>7.4</td>
</tr>
<tr>
<td>Coolguy</td>
<td>16.7</td>
<td>8.3</td>
<td>6.3</td>
</tr>
<tr>
<td>Paul Bachman</td>
<td>9.3</td>
<td>12.3</td>
<td>7.1</td>
</tr>
<tr>
<td>J.W.</td>
<td>14.5</td>
<td>10.2</td>
<td>8.1</td>
</tr>
</tbody>
</table>

Table 5.6 The DJs’ values for the three aggregate scales, Gender, Social Class, and Personality
also more susceptible to the influence of the phonological environment.

Phonetic studies of sibilant voicing show that there are sex-based differences. I discussed these differences to some extent in Chapter Two, based on the work of Titze (1989) and Koenig (2000). To briefly recap, phonation threshold (i.e., the ease with which a state of periodic vibration is obtained by the vocal folds) is directly related to the stiffness of the vocal folds and inversely related to vocal fold thickness. Males have thicker vocal folds, but they are less stiff, and therefore the phonation threshold for males is lower than for females. Koenig showed that males voice a greater portion of the fricative /h/ than females. Furthermore, the trend that women tend to devoice / not voice more often than men seems to hold for different languages (for French, see Temple 2000, reviewed in §2.2.2; for Japanese see Sugiyama 2004, reviewed in §3.5.1; for Spanish, see Wolf and Jiménez 1979 reviewed in §4.4.3; see also Sapir (reprinted in Mandelbaum 1958) for women devoicing final vowels in Yana).

Based on our hypothesis that sex-based anatomical variation forms a frame for gender-based variation, we predict that DJs with the higher perceived Gender ratings, as listed in Table 5.6, will also be the speakers with less devoicing in word-final sibilants.

5.4.2 Extracting the tokens

For each DJ, I extracted 200 tokens of word-final /z/ from one-syllable words. This large number of tokens was required because of the large number of environments considered (see next section). I did not select tokens by environment (for example, voicing of the following segment). I instead coded the environment factors in the token file. (The next section provides a list of environment factors.) Sibilant fricatives that were produced with an unusual vocal setting, such as a breathy vocal quality, laughing, etc., were excluded. Speech samples containing more than one speaker, miscellaneous background noise, or background music were also excluded. The fricatives were identified both auditorily and visually. The fricative boundaries were determined visually from the waveform and the spectrogram. The fricatives were primarily identified from the spectrograms. The visual feature of a fricative on a spectrogram is a concentration of energy above 3000 Hz. In the case of a transition from a vowel to a fricative, I also looked for a change in the pattern of the waveform. This is illustrated in Figure 5.2.
5.4.3 The independent linguistic variables

For each token, the following linguistic information was also recorded:

- the presence or absence of a morpheme boundary
- the phonological voicing of the following segment
- the sonority of the preceding segment
- the sonority of the following segment
- the duration of the fricative

The morpheme boundary was coded as either [yes] or [no], with the intention to recode by morpheme type if a significant effect was found. The following morphemes were coded: genitive case “Tom’s,” plural marking “books,” subject-verb case agreement “He
The voicing of the following segment was coded as either [voiced] or [voiceless]. A following affricate was coded by the first segment (the obstruent). Voiced hesitations such as *uh*, for example “*He’s uh he’s going*,” were included in the coding process. The sonority of the preceding and following segments were coded based on Selkirk’s (1984) Sonority Scale:

- 5 – vowels
- 4 – glides
- 3 – liquids
- 2 – nasals
- 1 – obstruents

Tokens that were followed by a pause or break in the speech that was greater than 25 milliseconds were excluded from the sonority correlations.

5.4.4 The dependent variable: Zero-crossing rate

Zero-crossing rate was first proposed over 35 years ago (Ito and Donaldson 1971) as a technique to recognize and distinguish speech sounds. Since then, this technique has gained popularity among applied linguistics applications, such as speech recognition (Lau and Chan 1985) and hearing aids (MacKinnon and Lee 1976).

Zero-crossing rate is the rate at which the speech signal crosses the 0 dB threshold in the temporal domain. Figure 5.3 illustrates the zero-crossing rates for a two fricatives produced by one of the DJs, one voiceless and one voiced. The voiceless fricative (Figure 5.3a) is essentially random noise, and therefore the speech signal crosses the 0 dB threshold at a very fast rate. The speech signal in Figure 5.3a crosses the 0 dB line 98 times in the ten milliseconds, and so the zero-crossing rate is 9800 crossings / second. Figure 5.3b illustrates the zero-crossing rate for the voiced alveolar fricative produced by the same DJ. The speech signal consists of two components, an underlying up-down pattern and a smaller random noise pattern overlaid on top. The up-down pattern is due to the periodic vibration of the vocal folds, the primary cue to phonological voicing in...
Figure 5.3a  Zero crossings for a voiceless alveolar fricative. The speech signal crosses the 0 dB threshold 98 times in ten milliseconds. Therefore, the zero-crossing rate is 9800 crossings per second.

Figure 5.3b  Zero crossings for a voiced alveolar fricative. The speech signal crosses the 0 dB threshold 18 times in ten milliseconds. Therefore, the zero-crossing rate is 1800 crossings per second.
fricatives. This underlying up-down pattern results in extended periods in which the speech signal, although it is still fluctuating randomly, is not crossing the 0 dB threshold. This greatly reduces the zero-crossing rate. The speech signal shown in Figure 5.3b crosses the 0 dB line eighteen times in ten milliseconds. The zero-crossing rate for the voiced fricative is therefore 1800 crossings / second, about 5.5 times lower than the rate of 9800 crossings / second for the voiceless fricative.

Zero-crossing rate is a gross measure of the voicing characteristics of the fricative. Because it looks at the entire fricative and calculates the average rate, the technique does not distinguish between a fully-voiced fricative that gradually changes to a fully-voiceless fricative, and a partially-voiced fricative that maintains a constant rate of voicing throughout the duration of the fricative. This, however, is not a problem. Both the gradual devoicing pattern and the constant partially-devoiced pattern, in comparison to a fully-voiced fricative, are representative of the feminine-engendered variant. Therefore, we are able to test our hypothesis with zero-crossing rate as the dependent variable. The acoustics program PRAAT version 4.4.2 (Boersma and Weenik 2006) was used to measure the zero-crossing rates. A detailed description and the code for the zero-crossing rate PRAAT script is provided in Appendix C. Table 5.7 lists the average zero-crossing rates for the eight DJs in order from highest to lowest. The zero-crossing rates vary from a low of 3902 crossings / second to a high of 8481 crossings / second. This wide range of crossing rates shows that there is a large amount of inter-speaker variability. I attempt to account for this inter-speaker variability with the various linguistic and social factors in the next section.

Table 5.7 Average zero-crossing rates for the eight DJs

<table>
<thead>
<tr>
<th>DJ</th>
<th>genre</th>
<th>Zero Xing Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Bachman</td>
<td>classical</td>
<td>8481</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>rock</td>
<td>7169</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>classical</td>
<td>6165</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>punk</td>
<td>6128</td>
</tr>
<tr>
<td>Country Dan</td>
<td>country</td>
<td>5692</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>pop</td>
<td>4571</td>
</tr>
<tr>
<td>J.W.</td>
<td>country</td>
<td>4035</td>
</tr>
<tr>
<td>Coolguy</td>
<td>metal</td>
<td>3902</td>
</tr>
</tbody>
</table>


5.4.5 The results

Let us begin by restating the hypotheses in terms of our variables. We predict the individual DJs will pattern according to their social gender with respect to devoicing of word-final sibilants. If this is the case, then the DJs’ Gender scores will show a negative correlation with zero-crossing rate, even after we have factored in the other linguistic and social variables. We also predict that the DJs who are perceived as working class will use more of the masculine variant. In other words, we predict that the DJs’ Social Class scores will show a positive correlation with zero-crossing rate.

Each of the linguistic variables fricative duration, morpheme boundary, preceding sonority, following voicing, and following sonority will be dealt with individually in turn, followed by the social variables. After examining each variable individually, interactions between the significant variables will be examined.

**fricative duration**

In order to test if fricative duration significantly correlates with zero-crossing rate, Pearson r correlations were run. Each speaker was run individually to avoid confounding speaker effects, such as the tendency for one speaker to speak slowly. Table 5.8 lists the token count, Pearson r statistic, and significance level for each for the DJs. Fricative

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>r</th>
<th>sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Dan</td>
<td>194</td>
<td>.284</td>
<td>.000</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>195</td>
<td>.648</td>
<td>.000</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>188</td>
<td>.462</td>
<td>.000</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>184</td>
<td>.455</td>
<td>.000</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>189</td>
<td>.397</td>
<td>.000</td>
</tr>
<tr>
<td>Coolguy</td>
<td>196</td>
<td>.021</td>
<td>.771</td>
</tr>
<tr>
<td>Paul Bachman</td>
<td>193</td>
<td>.308</td>
<td>.000</td>
</tr>
<tr>
<td>J.W.</td>
<td>190</td>
<td>.391</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 5.8 Pearson r correlations between zero-crossing rate and the fricative duration for each of the eight DJs.
duration significantly correlates with zero-crossing rate for seven of the eight DJs, with a longer duration corresponding to a higher zero-crossing rate. Therefore, in general fricative duration is related to zero-crossing rate.

**Table 5.9** Token counts and mean zero-crossing rates for two environments, part of a morpheme and not part of a morpheme, for the eight DJs

<table>
<thead>
<tr>
<th></th>
<th>Morpheme</th>
<th></th>
<th>Not Morpheme</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>mean</td>
<td>N</td>
<td>mean</td>
</tr>
<tr>
<td>Country Dan</td>
<td>117</td>
<td>5482</td>
<td>77</td>
<td>5958</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>109</td>
<td>4886</td>
<td>86</td>
<td>3998</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>106</td>
<td>6025</td>
<td>82</td>
<td>5902</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>100</td>
<td>7289</td>
<td>84</td>
<td>6681</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>95</td>
<td>6245</td>
<td>94</td>
<td>6003</td>
</tr>
<tr>
<td>Coolguy</td>
<td>109</td>
<td>3668</td>
<td>87</td>
<td>4009</td>
</tr>
<tr>
<td>Paul Bachman</td>
<td>93</td>
<td>8436</td>
<td>100</td>
<td>8439</td>
</tr>
<tr>
<td>J.W.</td>
<td>131</td>
<td>4000</td>
<td>59</td>
<td>3877</td>
</tr>
</tbody>
</table>

Table 5.9 lists token counts and mean zero-crossing rates by morpheme boundary environment for each of the DJs. A paired-samples t test comparing the means in the two environments shows that morpheme boundaries do not significantly correlate with zero-crossing rate ($t(7) = -0.91, p = .391$). Therefore zero-crossing rate cannot be considered to be related to presence of a morpheme boundary.

**sonority of the preceding segment**

Table 5.10 lists token counts and mean zero-crossing rates by sonority of the preceding segment. Glides were excludes as there were only six tokens for all eight speakers. A repeated measures one-way ANOVA with a Mauchly test of sphericity shows that the sonority of the preceding segment does not significantly influence zero-crossing rate ($F(3,21) = 2.16, p = .123$). Therefore, zero-crossing rate cannot be considered to be related to the sonority of the preceding segment.
Table 5.11 lists token counts and mean zero-crossing rates by voicing of the following segment for each of the DJs. A paired-samples $t$ test comparing the means in the two environments shows that voicing of the following segment significantly correlates with obstruents, nasals, liquids, and vowels.

<table>
<thead>
<tr>
<th></th>
<th>obstruents</th>
<th>nasals</th>
<th>liquids</th>
<th>vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>mean</td>
<td>N</td>
<td>mean</td>
</tr>
<tr>
<td>Country Dan</td>
<td>14</td>
<td>4115</td>
<td>22</td>
<td>7495</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>11</td>
<td>5468</td>
<td>33</td>
<td>5297</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>26</td>
<td>6874</td>
<td>20</td>
<td>6680</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>17</td>
<td>8095</td>
<td>28</td>
<td>7074</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>15</td>
<td>6897</td>
<td>28</td>
<td>6368</td>
</tr>
<tr>
<td>Coolguy</td>
<td>11</td>
<td>4038</td>
<td>30</td>
<td>4098</td>
</tr>
<tr>
<td>Paul Bachman</td>
<td>10</td>
<td>9671</td>
<td>39</td>
<td>8246</td>
</tr>
<tr>
<td>J.W.</td>
<td>9</td>
<td>4183</td>
<td>34</td>
<td>3311</td>
</tr>
</tbody>
</table>

Table 5.10  Token counts and mean zero-crossing rates by sonority of the preceding segment for the eight DJs

Table 5.11  Token counts and mean zero-crossing rates for two environments, before a voiceless segment and before a voiced segment, for the eight DJs

voicing of the following segment

Table 5.11 lists token counts and mean zero-crossing rates by voicing of the following segment for each of the DJs. A paired-samples $t$ test comparing the means in the two environments shows that voicing of the following segment significantly correlates with...
Therefore zero-crossing rate is related to the voicing of the following segment.

Table 5.12 lists token counts and mean zero-crossing rates before voiced segments, broken down by sonority of the segment. Only tokens before a voiced segment are included, as English lacks voiceless counterparts for the sonority levels beyond the lowest. Liquids were excluded due to low token counts (between two and eleven tokens per DJ). A repeated measures one-way ANOVA with a Mauchly test of sphericity shows that the sonority of the following segment does not significantly influence zero-crossing rate ($F(3,21) = .921, p = .448$). Therefore, zero-crossing rate cannot be considered to be related to the sonority of the following segment.

**The social variables**

Pearson $r$ correlations were run to test for a relationship between zero-crossing rate and each of the social variables. Table 5.13 lists the Pearson $r$ values and their significance levels for the correlations. Because we are running multiple correlations, we run the risk of coincidentally finding a significant correlation when there is not one. We reduce this risk with a Bonferroni correction to the alpha level, that is, we divide the alpha level by
the number of tests. Therefore only significance levels of 0.05 divided by five (I ran five tests), or 0.01, will be considered evidence for a significant relationship. With this adjustment in mind, we see that zero-crossing rate correlates significantly only with Gender, with a higher gender score corresponding to a lower zero-crossing rate.

interactions

Before we can confidently draw conclusions, we must check for interactions between Gender and the two linguistic variables that show significant correlations with zero-crossing rate, fricative duration and following voicing. It is possible that interactions between the linguistic variables and Gender are confounding the results. For example, it may be that the DJs with lower Gender values are also the DJs who produce longer duration fricatives due to a slower speech rate, and it is really speech rate that correlates with zero-crossing rate, not Gender. Or it may be that the DJs with lower Gender values are also the DJs who produce more tokens before a following voiceless segment, and it is the voicing of the following segment that correlates with zero-crossing rate, not Gender. If either of these scenarios is true, then we have a variable (either fricative duration or following voicing) lurking behind the correlation between Gender and zero-crossing rate, and the relationship between Gender and zero-crossing rate may be coincidental.

Let us begin with voicing of the following segment. Table 5.14 shows the breakdown of the tokens for each DJ by voicing of the following segment. The token counts exhibit a wide range of inter-speaker variation; just looking at the environment of
before a voiceless segment, we see that the counts range from as few as 42 tokens for Coolguy to as many as 76 tokens for Martin Goldsmith. I control for these differences in the number of tokens produced before voiced and voiceless segments by selecting the same number of tokens for each DJ. The ratio of voiced to voiceless tokens in Table 5.15 is roughly 2:1 for all of the DJs. In order to maintain this ratio, the token sets were reduced to the first 40 tokens before a voiceless segment and the first 80 tokens before a voiced segment. The revised zero-crossing rates for each of the DJs are listed in Table 5.15. Pearson $r$ correlations between the new rates and the social variables are listed in Table 5.16. \textit{Gender} still shows a significant correlation with zero-crossing rate after we have controlled for the voicing of the following segment.

I control for the effect of duration on the correlation between \textit{Gender} and zero-crossing rate with a hierarchical multiple regression analysis. Hierarchical multiple regression estimates each independent variable’s contribution to the prediction of the variance seen in the dependent variable, controlling for the other independent variables. The model is set up so that mean zero-crossing rate is the dependent variable, mean duration is the first predictor, and \textit{Gender} is the second predictor. This model will estimate how well \textit{Gender} predicts the variance in zero-crossing rate controlling for fricative duration. The regression model is run twice, once for tokens before a voiceless segment, and once for tokens before a voiced segment. Table 5.17 lists token counts, mean durations, mean zero-crossing rates and the \textit{Gender} values used in the analysis.
The results of the hierarchical multiple regression are presented in Table 5.18. The value in the $R^2$ Change column multiplied by 100 approximates the contribution the variable entered in that step makes to the overall prediction of the variance in zero-crossing rate. The last column in each table is the significance level for that variable’s contribution. For both environments, fricative duration does not significantly contribute to the prediction of the variance in zero-crossing rate. This lack of a significance is because much of the predictive power contained in fricative duration is also contained in the following voicing distinction. Comparing mean fricative duration across the two voicing environments in Table 5.17 shows that mean fricative duration is consistently longer before a following voiceless segment. Following voicing is also a primary determinant of zero-crossing rate (Table 5.11). Once we separate the tokens by following voicing environment, then fricative duration loses much of its power to predict zero-

<table>
<thead>
<tr>
<th>DJ</th>
<th>Old Zero Xing Rate</th>
<th>New Zero Xing Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Bachman</td>
<td>8481</td>
<td>8021.2</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>7169</td>
<td>6585.8</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>6165</td>
<td>5959.8</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>6128</td>
<td>5421.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DJ</th>
<th>Old Zero Xing Rate</th>
<th>New Zero Xing Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Dan</td>
<td>5692</td>
<td>5418.1</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>4571</td>
<td>3869.8</td>
</tr>
<tr>
<td>J.W.</td>
<td>4035</td>
<td>3760.5</td>
</tr>
<tr>
<td>Coolguy</td>
<td>3902</td>
<td>3935.8</td>
</tr>
</tbody>
</table>

Table 5.15 Average zero-crossing rates for each DJ controlling for following environment. For each DJ, only the first 40 tokens before a voiceless segment and the first 80 tokens before a voiced segment were used

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>r</th>
<th>sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>8</td>
<td>-.376</td>
<td>.359</td>
</tr>
<tr>
<td>Regional Accent</td>
<td>8</td>
<td>-.206</td>
<td>.625</td>
</tr>
<tr>
<td>Gender</td>
<td>8</td>
<td>-.825</td>
<td>.012</td>
</tr>
<tr>
<td>Social Class</td>
<td>8</td>
<td>.455</td>
<td>.257</td>
</tr>
<tr>
<td>Personality</td>
<td>8</td>
<td>.018</td>
<td>.967</td>
</tr>
</tbody>
</table>

Table 5.16 Pearson r correlations between zero-crossing rate and the social variables, controlling for token counts before following voiced and voiceless segments
The hierarchical regression model controls for the variables entered in previous steps. So in the case of step two, the $R^2$ Change column estimates the contribution of Gender to the prediction of the variance in zero-crossing rate once we have controlled for fricative duration. For both voicing environments, Gender significantly contributes to the prediction of the variance in zero-crossing rate. That Gender significantly correlates with zero-crossing rate supports Gender Hypothesis 2 for anatomically-grounded sex-based differences. The negative correlation between Gender and zero-crossing rate in Table 5.13 indicates that the DJs who value the facets of masculinity captured in the Gender scale, such as machismo, are the ones who are producing word-final fricatives with a greater degree of voicing. This difference in voicing is certainly not a result of sex-based anatomical differences in the vocal tract, as the DJs are all male. Rather, the tendency for men to voice consonants more frequently and to a greater extent leads to the association of voicing with masculinity. The DJs are using this association to index social gender.

Social Class Hypothesis 1 predicted a positive correlation between Social Class and zero-crossing rate. While the correlation was positive, it was not significant (Table 5.17 Mean zero-crossing rates and fricative durations before following voiceless and following voiced segments. Gender has also been added to the table for comparison)

<table>
<thead>
<tr>
<th>DJ</th>
<th>Gender</th>
<th>Before Voiceless</th>
<th></th>
<th>Before Voiced</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Dur.</td>
<td>Xing Rate</td>
<td>N</td>
</tr>
<tr>
<td>Country Dan</td>
<td>14.6</td>
<td>40</td>
<td>688 ms</td>
<td>7669</td>
<td>80</td>
</tr>
<tr>
<td>Ted Kelly</td>
<td>13.2</td>
<td>40</td>
<td>832 ms</td>
<td>5635</td>
<td>80</td>
</tr>
<tr>
<td>Lars Frederiksen</td>
<td>12.6</td>
<td>40</td>
<td>873 ms</td>
<td>7076</td>
<td>80</td>
</tr>
<tr>
<td>Chris Jericho</td>
<td>11.7</td>
<td>40</td>
<td>819 ms</td>
<td>7564</td>
<td>80</td>
</tr>
<tr>
<td>Martin Goldsmith</td>
<td>10.4</td>
<td>40</td>
<td>801 ms</td>
<td>6876</td>
<td>80</td>
</tr>
<tr>
<td>Coolguy</td>
<td>16.7</td>
<td>40</td>
<td>723 ms</td>
<td>5009</td>
<td>80</td>
</tr>
<tr>
<td>Paul Bachman</td>
<td>9.3</td>
<td>40</td>
<td>644 ms</td>
<td>8809</td>
<td>80</td>
</tr>
<tr>
<td>J.W.</td>
<td>14.5</td>
<td>40</td>
<td>753 ms</td>
<td>5748</td>
<td>80</td>
</tr>
</tbody>
</table>
The lack of significance does not necessarily imply the lack of a relationship between Social Class and zero-crossing rate. The second best correlation seen in Table 5.13, after Gender, is Social Class. These results neither support nor deny the claim that there is a relationship between Social Class and zero-crossing rate, but if there is a relationship the connection is a weak one.

### 5.5 Contextual and Inherent vowel length contrasts

#### 5.5.1 Vowel length contrasts as sex-based variation

Up to this point, I have demonstrated that anatomically-grounded sex-based differences in the devoicing of sibilants form a frame for the indexing of social gender. As I said earlier, this result is not novel. But it does lay down the foundation for the step into new territory — testing the hypothesis that not only anatomically-determined sex-based differences, but also behavioural sex-based differences create a frame for the indexing of social gender.

In order to test this hypothesis, I examine the contextual and inherent vowel length contrasts in the voices of the DJs. I reviewed a number of phonetic studies in Chapter Three that showed sex-based behavioural differences. These studies included two studies of vowel duration, one on sex-based differences in stressed and unstressed vowels in Swedish (Ericsdotter and Ericsson 2001, reviewed in §3.2), and the other on sex-based
differences in the production of phonemic short and long vowels in Creek (Johnson and Martin 2001, reviewed in §3.3). In both cases, women produce a sharper temporal distinction between the vowel categories than men. As I noted there, there is no anatomical difference that readily explains these temporal difference in the production of vowels. We predict that this behavioural sex-based difference will form a frame for social gender differences. Specifically, we predict that the DJs with the lower perceived Gender ratings (i.e., less macho-sounding) as listed in Table 5.6 will also be the speakers who show a sharper temporal distinction between the vowel duration categories.

The length of the vowel in English serves as a perceptual cue to a number of distinctions (Klatt 1976), such as:

- tense versus lax vowels
- the voicing of the following segment
- phrase-final versus non-final syllables
- stressed versus unstressed vowels

Tense vowels are inherently longer than lax vowels. Furthermore, vowels are longer before voiced segments than before voiceless segments. The contextual length pattern is overlaid on top of the inherent length pattern, so that, controlling for stress and sentence position, we end up with the four-way distinction shown in Table 5.19. In the narrow transcriptions, tense vowels are long and therefore are followed by a full colon. The following voiced consonant further lengthens the vowel, adding a half colon. We are interested in the contextually-determined duration differences (vowels before voiceless consonants versus vowels before voiced consonants) within each of the tense and lax vowel categories, and the inherent durational differences (tense vowels versus lax vowels) within each of the following contexts.

5.5.2 Extracting the tokens

The tokens were selected based on syllable structure, word class, and sentence accent, all of which have been shown to interact with syllable duration in English (Van Bergem 1993). Only monosyllabic words of CVC or CCVC syllable structure were considered, thus guaranteeing that every token had primary lexical stress. Of these, vowels in the
environment of a liquid or a glide were excluded, as the segment boundary between the
vowel and consonant proved to be too difficult to determine consistently from one token
to the next. Word class (lexical words versus non-lexical words) was controlled by only
selecting lexical words (nouns, verbs, adjectives); non-lexical items such as prepositions,
auxiliaries, etc., were excluded. Some examples of words that fit these criteria are name,
bit, noon, thing, beat, and spit. The sentential accent for each token was
impressionistically rated on a four-point scale: emphatic, heavily accented, lightly
accented, and unaccented. Accent level was determined both auditorily and by visually
comparing the vowel’s intensity peak to the intensity levels of the words in the rest of the
sentence. The heavily-accented tokens tended to be the word in the sentence with the
highest intensity level. Tokens that were judged to be either emphatic or unaccented were
excluded due to their infrequency. For each DJ, I extracted 100 vowel tokens that fit
these criteria. For each token, the duration of the vowel was measured by visually
locating the transitions on the spectrogram with PRAAT version 4.4.2 (Boersma and
Weenik 2006).

5.5.3 The dependent variables: Vowel length contrast ratios

The vowel tokens were divided into the four categories listed in Table 5.19. The
categories were further subdivided by accent level (heavily accent or lightly accented).
These three distinctions resulted in a total of eight categories: two vowel categories X
two contexts X two accent levels. I was not able to find an even number of tokens for
each category. The accented tense vowels before the voiceless consonant were the rarest.
In order to ensure adequate representation of each category, and to avoid over-
representation of one category, I chose a maximum of twenty tokens per category. The
minimum was seven tokens.

Two vowel ratios were calculated, a Contextual Vowel Contrast Ratio, and an
Inherent Vowel Contrast Ratio. The Contextual Vowel Contrast Ratio was calculated for
each of the DJs as follows. The mean vowel duration before a voiced consonant was
divided by the mean vowel duration before a voiceless consonant, holding the tense / lax
categories and the accent level constant, resulting in four ratios:

1. tense, heavily-accented vowels: before voiced C / before voiceless C
2. lax, heavily-accented vowels: before voiced C / before voiceless C
3. tense, lightly-accented vowels: before voiced C / before voiceless C
4. lax, lightly-accented vowels: before voiced C / before voiceless C

These ratios were summed, and four was subtracted from the value. If there is no difference between the contexts, then the ratio will be one, and the sum minus four will come to zero, so that a value close to zero represents a lack of a distinction between contextually long and contextually short vowels.

The Inherent Vowel Contrast Ratio was calculated in the same manner, but with the inherent and contextual categories reversed. Here are the four ratios:

1. heavily-accented vowels before voiceless C: lax vowel / tense vowel
2. heavily-accented vowels before voiced C: lax vowel / tense vowel
3. lightly-accented vowels before voiceless C: lax vowel / tense vowel
4. lightly-accented vowels before voiced C: lax vowel / tense vowel

Similar to the Contextual Vowel Contrast Ratios, the Inherent Vowel Contrast Ratios were also summed and four subtracted from the value.

5.5.4 The results

Table 5.20 lists the mean vowel durations and counts by category for each of the DJs. The means range from 52 milliseconds to 213 milliseconds. For every DJ, the heavily-accented vowels are longer than the lightly-accented vowels. The durational differences along the tense–lax dimension tend to be greater than durational differences along the
### DJ: Country Dan

<table>
<thead>
<tr>
<th>Heavily Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>166</td>
<td>143</td>
</tr>
<tr>
<td>Lax</td>
<td>125</td>
<td>125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>125</td>
<td>87</td>
</tr>
<tr>
<td>Lax</td>
<td>85</td>
<td>94</td>
</tr>
</tbody>
</table>

### DJ: Ted Kelly

<table>
<thead>
<tr>
<th>Heavily Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>167</td>
<td>137</td>
</tr>
<tr>
<td>Lax</td>
<td>127</td>
<td>92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>117</td>
<td>107</td>
</tr>
<tr>
<td>Lax</td>
<td>79</td>
<td>67</td>
</tr>
</tbody>
</table>

### DJ: Lars Frederiksen

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>177</td>
<td>112</td>
</tr>
<tr>
<td>Lax</td>
<td>92</td>
<td>79</td>
</tr>
</tbody>
</table>

### DJ: Chris Jericho

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>115</td>
<td>72</td>
</tr>
<tr>
<td>Lax</td>
<td>78</td>
<td>57</td>
</tr>
</tbody>
</table>

### DJ: Martin Goldsmith

<table>
<thead>
<tr>
<th>Heavily Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>198</td>
<td>136</td>
</tr>
<tr>
<td>Lax</td>
<td>112</td>
<td>80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>133</td>
<td>100</td>
</tr>
<tr>
<td>Lax</td>
<td>73</td>
<td>65</td>
</tr>
</tbody>
</table>

### DJ: Paul Bachman

<table>
<thead>
<tr>
<th>Heavily Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>185</td>
<td>141</td>
</tr>
<tr>
<td>Lax</td>
<td>105</td>
<td>87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>127</td>
<td>99</td>
</tr>
<tr>
<td>Lax</td>
<td>76</td>
<td>64</td>
</tr>
</tbody>
</table>

### DJ: J.W.

<table>
<thead>
<tr>
<th>Heavily Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>167</td>
<td>159</td>
</tr>
<tr>
<td>Lax</td>
<td>152</td>
<td>109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lightly Accented Syllables</th>
<th>voiced C</th>
<th>voiceless C</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Tense</td>
<td>121</td>
<td>129</td>
</tr>
<tr>
<td>Lax</td>
<td>87</td>
<td>74</td>
</tr>
</tbody>
</table>

Table 5.20 Average vowel durations for heavily-accented and lightly-accented vowels for the eight DJ
context dimension. In general, all of the DJs show the pattern seen in Table 5.19 for both accent conditions, although some of the lengths are reversed, particularly for the lightly-accented vowels. Six of the eight DJs show the four-way distinction in vowel length as transcribed in Table 5.19 for heavily-accented vowels, and four of the DJs show the four-way distinction for lightly-accented vowels, suggesting that the pattern begins to break down as the accent level reduces. But even then, the tense vowels are always longer than the lax vowels.

The shortest vowels of the DJs who show the greatest temporal distinctions are shorter than the DJs who show the smallest temporal distinctions. For example, the mean duration of the lightly-accented lax vowels before a voiceless consonant for Martin Goldsmith, who has the highest vowel contrast ratio, is 65 milliseconds. Compare this to the average duration of 94 milliseconds for the same category for Country Dan, who has the lowest vowel contrast ratio. Now compare their durations for the longest category, the heavily-accented tense vowel before a voiced consonant. Martin Goldsmith shows a value of 198 milliseconds, notably longer than Country Dan’s average duration of 166 milliseconds. In other words, Martin Goldsmith’s long vowels are longer than Country Dan’s, but Martin Goldsmith’s short vowels are shorter than Country Dan’s. Furthermore, this pattern holds for the other DJs as well. This pattern shows that the differences in vowel durations are not simply a reflection of different speech rates; if they were then we would expect that the DJs with the longest long vowels would also have the longest short vowels. Rather, the vowel ratios are reflection of differences in the phonetic distinctiveness of the phonological contrast length contrasts.

Table 5.21 lists the Contextual Vowel Contrast Ratios and the Inherent Vowel Contrast Ratios for each of the DJs. The Contextual Vowel Contrast Ratios for the eight DJs ranged from 0.50 to 1.30, while the Inherent Vowel Contrast Ratios ranged from 0.87 to 2.83. Figures 5.4 and 5.5 are visual displays of the same information. Figure 5.4 shows the average vowel durations before a voiced and voiceless consonant for each of the eight DJs for the two accent conditions. Figure 5.5 shows the average vowel durations for tense and lax vowels for each of the eight DJs for the two accent conditions. The Contextual Vowel Contrast Ratios represent the different in heights between the pairs of light-coloured and dark-coloured bars in Figure 5.4. Similarly, the different in
heights between the pairs of bars in Figure 5.5 represent the Inherent Vowel Contrast Ratios.

Pearson $r$ correlations were used to determine if any of the independent social variables, Gender, Social Class, Personality, Regional Accent, and Age, significantly correlate with the vowel context ratios. Again, we use a Bonferroni correction to the alpha level and one-sided correlations for Gender and Social Class. Table 5.22 lists the Pearson $r$ correlations and their significance levels.

Of the correlations with the Contextual Vowel Contrast Ratios, none are significant. Gender shows the strongest correlation ($r(6) = -.747, p = .033$), but after the Bonferroni correction, it is slightly beyond marginal. Gender does show a significant correlation with the Inherent Vowel Contrast Ratios ($r(6) = -.880, p = .004$). Evidently, inherent vowel length is a better conveyer of social information than contextual vowel length. This is not surprising. The phonological contrast between a tense vowel and lax vowel is that between a long vowel and a short vowel, whereas the phonological contrast between a contextually long vowel and short vowel is that between a half-long vowel and a short vowel (see Table 5.19). Furthermore, the temporal distinction between a contextually long vowel and short vowel breaks down for some of the DJs in the lightly-accented condition, whereas the distinction between the inherently long and short vowels is still robust in the lightly-accented condition. So in general, inherent vowel length is a better conveyer of both linguistic information and social information.

These results support the hypothesis that behavioural sex-based differences create a frame for the indexing of social gender. It seems that the pattern that argued for in
Chapter Three — women produce greater temporal distinctions for a given temporal-based phonological contrast than men — is picked up on by the DJs, and used to index their gender roles so that the DJs who project masculinity more are also those DJs who produce less of a temporal distinction between contextually long and short vowels.

An examination of the $r$ values in Table 5.22 again shows that Social Class has the second strongest correlation with the vowel contrast ratios, after Gender, although the correlations are not significant. This pattern was also seen for the correlations between the social variables and zero-crossing rate, so we have now seen the same order of strengths ($Gender$ the strongest, $Social Class$ the second strongest) for three variables. Collectively, these results are suggestive of a weak relationship between $Social Class$ and the vowel contrast ratios, supporting the $Social Class Hypothesis 1$. 

Figure 5.4  Average vowel durations before a voiced and voiceless consonants for the eight DJs in two accent environments, heavy (above) and light (below)
Chapter Summary

The objective of this chapter was to demonstrate that phonetic-level sex-based differences in our speech create a frame for indexing social gender. I began with an example of an anatomically-motivated sex-based difference, the devoicing of word-final sibilants. Physiological differences between the vocal folds of men and women may result in men voicing more than women. This pattern creates a frame for the indexing of social gender so that speakers who value the macho aspects of masculinity more tend to voice their word-final sibilants more. We saw this for the eight DJs. I then provided an examination of two behaviourally-motivated sex-based differences, contextual and inherent vowel length contrasts. Behavioural differences between men and women result in...
in women producing sharper phonetic distinctions for phonological contrasts such as vowel length. This pattern also creates a frame for the indexing of social gender so that speakers who value the macho aspects of masculinity more tend to produce less of temporal distinction between contextual and inherent long and short vowels. This pattern was also demonstrated with the speech of the eight DJs.

There is an interesting paradox in the behaviour of the DJs with respect to the maintenance of the phonological contrasts examined in this chapter. To quickly recap, the primary perceptual cue to the phonological contrast between voiced and voiceless fricatives is periodic vibration of the vocal folds. Therefore, devoicing results in a reduction of the phonetic distinctiveness of the phonologically voiced fricatives. One other phonological contrast examined was contextual vowel duration. The duration of the preceding vowel is an important cue to the perception of the phonological voicing of a word-final stop (Nittrouer 2004, Wang and Wu 2001). Therefore a reduction in the temporal differences between contextually-determined long and short results in a loss of the phonetic distinctiveness of the voicing contrast in word-final stops. So we should expect that the DJs who value the maintenance of phonetic distinctiveness should both voice the word-final sibilants and produce large temporal differences between contextually-determined long and short vowels. But in fact the DJs who maintain the phonetic distinctiveness in word-final sibilants are also the DJs who reduce the phonetic

<table>
<thead>
<tr>
<th>Contextual Vowel Length Ratio</th>
<th>N</th>
<th>r</th>
<th>sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>8</td>
<td>-0.271</td>
<td>0.516</td>
</tr>
<tr>
<td>Regional Accent</td>
<td>8</td>
<td>-0.501</td>
<td>0.206</td>
</tr>
<tr>
<td>Gender</td>
<td>8</td>
<td>-0.747</td>
<td>0.033</td>
</tr>
<tr>
<td>Social Class</td>
<td>8</td>
<td>0.606</td>
<td>0.111</td>
</tr>
<tr>
<td>Personality</td>
<td>8</td>
<td>-0.415</td>
<td>0.307</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inherent Vowel Length Ratio</th>
<th>N</th>
<th>r</th>
<th>sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>8</td>
<td>-0.145</td>
<td>0.732</td>
</tr>
<tr>
<td>Regional Accent</td>
<td>8</td>
<td>-0.366</td>
<td>0.373</td>
</tr>
<tr>
<td>Gender</td>
<td>8</td>
<td>-0.880</td>
<td>0.004</td>
</tr>
<tr>
<td>Social Class</td>
<td>8</td>
<td>0.692</td>
<td>0.057</td>
</tr>
<tr>
<td>Personality</td>
<td>8</td>
<td>-0.386</td>
<td>0.344</td>
</tr>
</tbody>
</table>

Table 5.22  Pearson $r$ correlations between Contextual Vowel Contrast Ratios and the independent and social variables
The answer to this puzzle lies in the indexation of social gender. Both the voicing of word-final sibilants and the reduction of the temporal distinction between contextually-determined long and short vowels are masculine engendered variants. Thus, the seemingly contradictory maintenance of phonological contrasts is consistent with the way the DJs are indexing of social gender.

Although periodic voicing is the primary cue to the phonological voicing of word-final fricatives, it is not the only cue, one of the others being the duration of the preceding vowel. So the DJs who devoice their word-final sibilants can still maintain the phonetic distinctiveness of the phonological voicing contrast in voicing by using other perceptual cues. And we predict they should be doing just that, since the DJs who devoice word-final sibilants are also the DJs who value the maintenance of phonological contrasts, as seen in their greater temporal distinctions for both contextual and inherent long and short vowels. Thus, average zero-crossing rate controlling for voicing and the Contextual Vowel Duration Ratios should show a positive correlation, and they do, albeit, a very weak one ($r(6) = .555$, $p = .144$). So it appears that the loss of phonetic
distinctiveness of the voicing of word-final sibilants resulting from a lack of periodicity of the vocal folds is compensated for by a sharper distinction made in the length of the preceding vowel.

This example shows the complexity of the interconnectivity between the subphonemic variation of different phonemes and our social gender. Because every phoneme in every language is involved in a phonological contrast, they are all connected to our social gender, as illustrated in Figure 5.7. Woven into that web of interconnectivity is also the anatomically-motivated engendered variation. In the next chapter I will look at one implication of this interconnectivity for our understanding of language change — the role of sex-based behavioural differences in vowel mergers.
Phonetic Distinctiveness in North American English Vowels

The range of free variation is large, which makes it difficult to develop descriptive models and quantitative rules.

Fant et al. 1991:528

6.1 Chapter objective

The previous chapter demonstrated that behaviourally-motivated sex-based differences created a frame for the indexing of gender. In that chapter, we saw that the phonetic distinctiveness of the contextually long and short vowels correlated with social gender. The objective of this chapter is to further demonstrate that phonetic distinctiveness is able to take on social significance, and more importantly, that phonetic distinctiveness plays a role in vowel mergers in North American English. I begin this chapter with the observation introduced in Chapter Three that men are more closed-mouth than women (Labov 1972:304). This is reflected in the smaller vowel space for men as compared to women, even after normalization. My objective is to demonstrate that this sex-based variation in the pronunciation of vowels is engendered, i.e., able to take on social significance, and that it shows several of the characteristics of engendered variation.

The last chapter looked at the interaction between phonetic distinctiveness and fine-grained perceptions of social gender for a small number of speakers. In this chapter, I switch to a much larger scale, with data from over 400 speakers. With so many speakers, we must use biological sex instead of fine-grained social gender. But we now have a number of speakers from a wide range of social class backgrounds. The primary objective of the first half of this chapter is to demonstrate that phonetic distinctiveness correlates with social class. The specific way the hypotheses are tested is explained in
§6.3.2, after the vowel distinctiveness metrics are introduced.

The second half of the chapter demonstrates the implications phonetic distinctiveness has for our understanding of sex differences in vowel mergers. The vowel merger process results in a loss of the phonetic distinctiveness vowel of the vowels involved. Because vowel mergers are current changes in progress, they allow us to test one of the language change hypotheses introduced in §4.4, that men lead linguistic changes that involve the loss of phonetic-level distinctiveness.

The data for both of these investigations comes from the *Atlas of North American English* by William Labov, Sharon Ash, and Charles Boberg (2006). The atlas is introduced in detail in the following section.

### 6.2 The Atlas of North American English

The *Atlas of North American English* (hereafter ANAE) is a detailed examination of the variation in vowel systems throughout English-speaking Canada and the United States. The atlas project team was headed by William Labov, Sharon Ash, and Charles Boberg. The data were gathered between 1992 and 2000 by telephone survey, and the atlas itself was published in 2006. The speakers were selected from telephone listings. Last names that occurred in clusters were given preference, as this maximized the likelihood of reaching a speaker who is a native of their place of residence. Altogether, the project team interviewed 805 speakers of which 762 were selected as satisfying the project’s definition of a local speaker (i.e., an individual who was born and grew up close to where they are currently residing). See Labov et al. (2006; chapter 4) for further details.

#### 6.2.1 The interview

All of the interviews were conducted over the telephone and recorded by using a telephone signal splitting device. The earlier interviews were recorded on a Nagra IV, a Nagra E, or a Tandberg Model 9021 reel-to-reel tape recorder. Later interviews were recorded on digital cassette tapes using a SONY TCD-D8 DAT recorder.

The complete interview was carried out in two stages. The first stage consisted of establishing contact with the speaker by telephone, and then interviewing him or her. The interview began by eliciting background information from the speaker, such as place of birth, residence history, father’s and mother’s place of birth, and languages spoken. This
was followed by a period of spontaneous conversation involving such topics as recent developments in the city, the state of the downtown area, and travel outside the city. If a topic of special interest to the speaker was found, it was developed to the fullest extent possible.

The spontaneous speech portion of the interview was followed by the elicitation of word lists, minimal pairs, spontaneous pronunciations of key lexical items, and grammatical acceptability judgments. The word lists consisted of words that did not require reading: counting, days of the week, articles of clothing, breakfast foods, etc. Minimal pairs such as hot/caught were elicited by first prompting the speaker for the first half of the pair with a question such as, “What is the opposite of cold?” This was then followed by the request for a judgement on how well the two words rhymed, in terms of “same” or “different.” Grammatical acceptability judgments were elicited for several sentences that the interviewer read to the speaker. The speaker then rated each sentence using a three point scale: 1 “could say yourself,” 2 “heard but wouldn’t say,” and 3 “never heard.” Spontaneous productions of key lexical items were elicited with the Semantic Differential technique (Labov 1984). This entailed asking for the difference between similar items, such as a bunk and a cot, where one of the two items was the key lexical item being sought. This resulted in multiple highly stressed productions of the key lexical item. The first stage of the interview closed with the elicitation of more general background information, such as occupation, education, and national ancestry. This was followed by a request that the speaker be interviewed a second time in the form of reading words from a word list that was to be provided to the speaker in the mail.

The second stage of the interview consisted of the speaker reading the full page word list, which was designed to further investigate lexical distributions of vowel classes. This was followed by a more detailed investigation of patterns of travel, friendships, kinship, and relations with other cities of interest.

6.2.2 Impressionistic coding

The ANAE provides two main types of data on the vowels of North American English: the presence or absence of mergers in the speakers vowel system, and the precise location of vowels in the form of formant measurements. The mergers data are in the form of impressionistic judgments of the minimal pairs by the both the speaker and the
analyst. The speaker judged minimal pairs produced by the interviewer as either “same” (i.e., merged), “close,” or “different.” An analyst then judged the productions of minimal pair words read by the speaker as either “the same,” “phonetically close,” or “clearly distinguishable.” The addition of the analyst’s judgments to those of the speakers allowed for an investigation of whether production or perception leads mergers by carefully comparing the two sets of judgments.

6.2.3 Formant analysis

The atlas project team carried out first and second formant analyses for many, but not all of the speakers. Altogether, formant measurements were made for 439 speakers. Formant analysis was carried out with the Computerized Speech Lab (CSL) program developed by Kay Electronics. The interview tapes were digitized at a sampling rate of 11,000 Hz using the CSL equipment.

Because the interviews were conducted over the telephone, the frequency range of the speech signal was limited to that transferred over telephone lines, about 300 to 3,000 Hz. This resulted in a reduction of the sound quality of the recordings. Furthermore, the speech signal was often accompanied by background or mechanical noise. However, the researchers found that quality of the speech signal was still high enough to carry out acoustic analysis with a satisfactory degree of confidence and reliability. Although the recordings were noisy, they still resulted in spectrograms with clearly interpretable formant structures.

Tokens of vowel productions were extracted from the spontaneous speech, minimal pairs, word lists, and the productions of key lexical items. Only highly stressed tokens were selected to avoid the confounding effects of vowel reduction in non-stressed environments. Much of the data collected came from spontaneous speech. The researchers were therefore unable to obtain identical tokens for each speaker. The researchers instead attempted to measure a similar balance of vowels for each speaker. This resulted, for the most part, in between five and ten tokens of each vowel, although there are a limited number of cases where no tokens were found. Measurements for each vowel were limited to at most ten tokens to avoid over-representation by one or two frequently produced vowels. These methods resulted in approximately 300 tokens per speaker, with some speakers having as few as 200 tokens, while others speakers had as
many as 400 or 500 tokens. Altogether, approximately 134,000 vowel tokens were measured. The average values for each formant for each vowel are published in a Microsoft Excel spreadsheet format on the multimedia CD that accompanies the print version of the atlas.

Once a suitable token was identified two formant measurements were made, one for the first formant and one for the second formant. In order to pick the point of measurement for the vowel formants, the main trajectory of the tongue was first identified. In vowels that involve first a lowering and then a raising of the tongue, such as most short vowels and many of the long upgliding vowels, F1 showed a clear inflection (i.e., change from rising to falling), indicating the lowest point of the tongue. This point was used for both formant measurements. The other main trajectory of the tongue is a backward-forward motion. This is seen in ingliding vowels, and is reflected in spectrograms by an inflection in the second formant. For these vowels the inflection point of the second formant was used as the point of measurement for both formants.

6.2.4 The social characteristics of the speakers

Each of the speakers was coded for several social characteristics, of which I make use of three: sex, socio-economic index, and dialect region. The other characteristics are ethnicity and city of residence. Sex is either male or female. Socio-economic index (SEI) scores are assigned by the researchers from published scores for the various occupations (Duncan 1961, Nakao and Treas 1989, Nakao and Treas 1992). These scores reflect not only the education and income levels of the occupation, but also the prestige associated with the various occupational titles, and therefore are an improvement over just the occupation itself. The scores ranged from 18 at the lower end through to 96 at the upper end. Several of the speakers did not provide adequate occupational information to assign a SEI score. Either their job description was too vague, such as “I work in an office,” or their job description did not match one of the published occupations. SEI scores were not assigned to these speakers. Women who reported themselves as homemakers were assigned a SEI score according to their husband’s occupation. Youth who had yet to enter the job market on a full-time basis were assigned a SEI score based on the primary income of the household. College students are in a transitional state, and may be upwardly mobile. They therefore were left unclassified. All together, 131 of the subjects
were not classified for SEI. Of the 439 speakers with acoustic data, the SEI index ranges from 22 to 96, and 75 speakers were not assigned a SEI score.

Eight major dialect regions along with several minor dialects within those regions were identified within the English of North America. These are listed in Table 6.1. Their identification was based on their participation or lack of participation in the various vowel shifts, mergers and changes that are taking place in North American English. Table 6.1 lists the various changes in progress that were used to identify the dialects.

**6.3 Vowel distinctiveness and engendered variation**

**6.3.1 The metrics**

In order to test the hypothesis that vowel distinctiveness can take on social significance, we need a measure of how dispersed the vowels are. It is possible to create a metric of vowel distinctiveness based on each vowel’s distance from the average of all vowel formants (Bradlow et al. 1996, Pierrehumbert et al. 2004). However, this approach is not sensitive to regional variation in vowel distinctiveness within the vowel space. I will instead use two simple metrics of vowel distinctiveness: the distance between /iy/ and /æ/, or Front Vowel Distinctiveness; and the distance between /uw/ and the low back vowel /o/, or Back Vowel Distinctiveness. There are a number of vowel changes currently taking place in North American English. Some of the changes involve the front vowels while others involve the back vowels. By keeping the measure of vowel distinctiveness to within either the front or the back region of the vowel space, we gain more control over the interaction between vowel distinctiveness and vowel changes. Including vowels that are undergoing sound changes introduces the possibility of confounding factors, and therefore must be avoided. If we include vowel productions from speakers involved in vowel sound changes such as the Northern Cities Vowel Shift, then we will not know if any social differences found are due to differences in vowel distinctiveness of the different social groups, or due to different levels of participation in the sound change by the different social groups. To include more than these four vowels runs the risk of confounding factors. As it is, even with these four vowels we must

---

10 Throughout this chapter I use the phonemic notation of ANAE. Square brackets indicate broad IPA transcription.
eliminate much of the available data in order to guarantee that only the most conservative productions of the vowels are included in the data set. For the back vowels there is only one dialect region in which the two vowels are stable enough to include in our investigation. If we create metrics of vowel distinctiveness that use more than four vowels, there is not a single dialect region in North America with vowels stable enough to include in the data set.

<table>
<thead>
<tr>
<th>Region / Dialect</th>
<th>Defined by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>low back merger, fronting of /uw/ but not /ow/</td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>Canadian Raising, fronting of /ahr/</td>
</tr>
<tr>
<td>Other Canada</td>
<td>Canadian Shift, Canadian Raising</td>
</tr>
<tr>
<td>Eastern New England</td>
<td>vocalization of /r/, short a nasal system</td>
</tr>
<tr>
<td>Boston</td>
<td>fronting of /ah/ distinct from /o/, low back merger</td>
</tr>
<tr>
<td>Providence</td>
<td>conservative /uw/ and /ow/, no low back merger</td>
</tr>
<tr>
<td>Midland</td>
<td>transitional low back merger, fronting of /ow/</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>short a system</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>glide deletion of /aw/</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>no vocalization of /r/, split of short a</td>
</tr>
<tr>
<td>North</td>
<td>conservative /ow/, no low back merger</td>
</tr>
<tr>
<td>Inland North</td>
<td>Northern Cities Shift</td>
</tr>
<tr>
<td>St. Louis</td>
<td>/ah/ and /oh/ merged</td>
</tr>
<tr>
<td>Western New England</td>
<td>less advanced Northern Cities Shift</td>
</tr>
<tr>
<td>North Central</td>
<td>low back merger, conservative /uw/ and /ow/, no Canadian Raising</td>
</tr>
<tr>
<td>New York City</td>
<td>vocalization of /r/, split of short a</td>
</tr>
<tr>
<td>South</td>
<td>glide deletion of /ay/</td>
</tr>
<tr>
<td>Inland South</td>
<td>Southern Vowel Shift</td>
</tr>
<tr>
<td>Texas South</td>
<td>Southern Vowel Shift</td>
</tr>
<tr>
<td>Southeast</td>
<td>fronting of /uw/, /ow/</td>
</tr>
<tr>
<td>Charleston</td>
<td>monophthonal /e:/ and /o:/, r-less</td>
</tr>
<tr>
<td>Florida</td>
<td>fronting of /uw/ and /ow/</td>
</tr>
<tr>
<td>Transitional</td>
<td>low back merger, fronting of /uw/ but not /ow/, no Canadian Raising</td>
</tr>
<tr>
<td>Western Pennsylvania</td>
<td>low back merger</td>
</tr>
</tbody>
</table>

Table 6.1 List of major and minor dialect regions identified in the *Atlas of North American English*. From Labov et al. (2006:146)
Although the regions included in the data set for the back vowels are limited, back vowel distinctiveness is still a useful criterion. By testing the same metric twice, once on the front vowels and once on the back vowels, we build in redundancy. This way if similar results are seen for both metrics, we can be confident that the results reflect a real-world phenomenon and not a statistical anomaly.

The Front Vowel Distinctiveness (FVD) metric is the phonetic distance between /æ/ and /iy/, while the Back Vowel Distinctiveness (BVD) metric is the phonetic distance between /uw/ and /o/. The relevance of the distance between adjacent vowels to the quality of the vowels themselves was first proposed by Lindblom (Liljencrants and Lindblom 1972, Lindblom and Sundberg 1969, Lindblom and Sundberg 1971). These researchers analogize vowel space with an electric field. They equate vowels to particles with an equal charge. Like particles in an electric field, vowels move apart from each other until they cannot disperse any further, and an equilibrium is reached (Liljencrants and Lindblom 1972:841). The researchers claim that similar to particles in an electric field, the influence a vowel excerpts on other vowels in the system decreases exponentially with distance. They build these assumptions into a mathematical model that they use to predict ideal vowel locations in an F1 by F2 by F3 space for a given number of vowels. I build the assumption that vowel influence decreases exponentially with distance into the measures by using two measures, one for front vowels, and one for back vowels.

I calculate the distance between two vowels in a linear fashion by treating the first and second formant measurements as points in Euclidian space, and then calculating the distance between the points (Yang 1996). For example, the distance between /iy/ and /æ/ is:

$$d = \sqrt{(F1_a - F1_i)^2 + (F2_a - F2_i)^2}$$

The subscripts indicate the vowel. For example, F1_a is the average formant value for the first formant of the low front vowel /æ/, as reported in ANAE.

The formant values published in the ANAE are in Hertz, and are not normalized. Labov provides the formula that they used to normalize the vowels. I do not normalize the data, as the objective of normalization is to remove inter-speaker differences in vowel
space and allow for a direct comparison of vowel positions. I, however, am examining
differences in vowel space and am not interested in the location of the same vowel in
different speakers, but rather their position relative to other vowels in the same vowel
system. Because normalization alters the shape of the vowel system by stretching or
shrinking the overall vowel space with a mathematical transformation, it destroys any
possibility of a comparison of distinctiveness differences between speakers.\textsuperscript{11}

I do, however, convert all of the formant values from Hertz to mels. The mel scale
is non-linear, and reflects more accurately how the human ear actually perceives pitch.
For example, a tone that is perceived to be twice as high in frequency as another tone is
double the number of mels. This is not true for the Hertz scale which does not reflect
actual human perception of tones. Because we are concerned with the way people
perceive differences in the voices of males and females, and then use those perceived
differences in the construction of social gender, it is important to use a psycho-acoustic
scale such as mels or Bark. As well, Lindblom and the other researchers mentioned
above found that their mathematical model of vowel space greatly improved when they
replaced the linear Hertz scale with the non-linear mel scale (Liljencrants and Lindblom
1972).

6.3.2 Hypotheses tested

The objective of the first half of the chapter is to test one of the gender hypotheses
introduced at the end of Chapter Three, and the two social class hypotheses introduced in
Chapter Four. Each of the hypotheses is repeated here, followed by a rephrasing in terms
of FVD and BVD.

Gender Hypothesis 1 (page 84)

The phonetic distinctiveness of the speech of women is greater than that of
the speech of men. In other words, women will show a greater phonetic
distance between /iy/ and /æ/, and between /uw/ and /o/. This will be
reflected in greater average FVD and BVD values for the women than the
men.

\textsuperscript{11} The lack of normalization means that the data set used here is not identical to the data set used in the
ANAE. As well, a theoretically interesting question remains: Do listeners normalize out the sex
differences seen here (Jane Stuart-Smith, personal communication)? If they do, a careful comparison of
the normalized data with the non-normalized may lead to a better understanding of why the perception
of mergers leads the production of mergers (see §6.4.3 for discussion). This work remains for the future.
Social Class Hypothesis 1 (page 92)
Working-class speakers use more masculine variants than middle-class speakers, while middle-class speakers use more feminine variants than working-class speakers. A greater average FVD and BVD is considered the feminine variant. Therefore, this hypothesis will be reflected in a correlation between vowel distinctiveness and social class for both front and back vowels.

Social Class Hypothesis 2 (page 93)
Working-class speakers show greater sex differentiation for both masculine and feminine engendered variants than middle-class speakers. This will be reflected in greater sex differences between the vowel distinctiveness metrics for the working-class speakers than for the middle-class speakers.

6.3.3 Preliminary results for FVD
Let us begin by comparing the average FVD by sex for each of the dialects. We expect that in every case, the females should show a greater average FVD than the males. Table 6.2 lists the average FVD in mels by sex for each of the dialect regions identified in the ANAE. Not every region has data for both males and females, and one region, North Central, lacks both. The overall trend is for females to have a larger FVD than males, 195.7 mels for the females compared to 183.7 mels for the males. The difference between the two means is statistically significant ($t(436) = 2.44, p = 0.015$). This corroboration of Gender Hypothesis 1, that women will show a greater phonetic distance between /æ/ and /iy/ than men, is not surprising given the literature reviewed in Chapter Three. At this point, there is no indication that a closed-mouth articulatory setting can take on social significance because the sex differences may be a result of differences in vocal tract length. In order to ascribe any social significance to a closed-mouth articulatory setting, we must look for a correlation with a social characteristic of the speakers who is not physically determined. As mentioned earlier, the attribute I will use for this purpose is social class.

Pearson $r$ correlations were run on the males and females separately to test for correlations between FVD and SEI. Several of the speakers lacked a SEI score due to insufficient information elicited during the interview. These speakers were removed from the analysis, leaving 147 males and 215 females. The correlation between FVD and SEI was significant neither for the males ($r(145) = .113, p = .174$), nor the females ($r(213) = .031, p = .656$). These results suggest that vowel distinctiveness does not convey social
Table 6.2 The average distance between /æ/ and /iy/ in mels for each dialect region by sex

information about the speaker. However, it turns out that ongoing language changes confound our results. The following section reanalyzes the data without the confounding regions.
6.3.4 Reanalysis of FVD without the confounding regions

Although most of the dialect regions show the general trend of women having greater FVD vowels than men, there are five dialect regions that show the reverse trend: Other Canada, Boston, St. Louis, Inland North, and Other Western Pennsylvania. In Boston, St. Louis, and Other Western Pennsylvania, the number of speakers is too low to draw conclusions, but in the other two regions the higher FVD values for the men suggests that they are making a sharper phonetic distinction between /æ/ and /iy/ than the women.

Both /æ/ and /iy/ are involved in various vowel shifts across the different regions. Labov et al. (2006: Part C) examine in detail regional variation of F1 and F2 across North America through a series of dialect maps that indicate the relative height and fronting / backing of the vowels across North America. Below each map is a brief explanation of the variation seen in the map. The variation reported for the first and second formants of /æ/ and /iy/ in the atlas can be summarized as follows (page references refer to the ANAE maps and their explanations):

the relative height of /æ/ (p. 82)

The Inland North region, extending from south-eastern Wisconsin to New York state, shows a large number of extremely high variants. This is the first stage of the Northern Cities Shift. New York and the Mid-Atlantic show a clustering of extremely low variants. This reflects the splitting of /æ/ into tense and lax variants, of which the laxed variant (the variant used for comparison here) is not raised. Other Canada shows a consistently low variant, while Atlantic Canada shows a tendency of raising, although not to the extent of the Inland North area.

the relative fronting and backing of /æ/ (p. 83)

The variation in the fronting and backing of /æ/ in many ways mirrors the variation seen in the height of /æ/. The Inland North region stands out as using almost exclusively an extremely fronted variant. However, unlike the raising, the extreme fronting extends as far as St. Louis and Kansas City. Other Canada displays predominantly the extremely low variant. This reflects the Canadian Shift in this region, in which the /æ/ vowel is extremely backed and slightly lowered.

the relative height of /iy/ (p. 92)
The South and southern Mid-Atlantic regions show a consistently lower height than the other regions. This is part of the third stage of the Southern Shift. Both regions of Canada are distinguished by a consistently high variant.

Consistent with the relative height of /iy/, the South and Mid-Atlantic regions show a consistently more backed variant than the other regions. The other regions show a variant that is higher and more fronted. Both regions of Canada almost consistently show an extremely fronted variant.

In their explanation of the variation in /i/ and /æ/ in the ANAE summarized here, Labov et al. identify four language changes that alter the distance between /iy/ and /æ/: the Northern Cities Vowel Shift (Inland North, St. Louis, Western New England), the Canadian Shift (Other Canada), the Southern Shift (Inland South, Texas South, Other South), and the split of /æ/ into tense and lax variants (Mid-Atlantic, New York City). Of these, the three chain shifts are problematic in that they introduce confounding variables into the analysis. Even if we see a clear correlation between social class and FVD among these groups of speakers, we still cannot draw any conclusions because it may well be that it is the vowel shift itself that has taken on social significance, and not vowel distinctiveness. The speakers from these regions (Inland North, St. Louis, Western New England, Other Canada, Inland South, Texas South, Other South) need to be removed from the analysis.

The split of /æ/ into tense and lax variants, on the other hand, is not a concern because it is the tense, raised variant that has taken on (negative) social significance, and when attention is given to speech, it is corrected to the non-raised, lax variant (Labov et al. 2006:173). As such, the production of /æ/ is stable in these dialects.

After the problematic regions are removed, 250 speakers remain, of which 48 lack SEI scores, and therefore also need to be removed. Furthermore, the removal of the problematic regions greatly reduced the variability in the data. A recheck of the distribution of the speakers by sex revealed that two of the women are outliers (FVD
values of 334 and 98), and one of the women is an extreme outlier (FVD of 57). The extreme outlier was removed, but the other two outliers were left in.

We now need to re-examine the relationship between social class and FVD for the remaining 201 speakers. The correlation between FVD and SEI for these speakers is marginally significant ($r(199) = .128$, $p = .071$). When we break down the data by sex, then the correlations for both the 81 males ($r(79) = .202$, $p = .070$) and the 119 females ($r(118) = .329$, $p = .090$) are marginally significant. The positive $r$ values for both of the correlations indicate that as social class increases, the distance between the two front vowels increases. Figure 6.1 shows the scatterplots of FVD and SEI for the males and females.

These results give some support to Social Class Hypothesis 1, that speakers from the lower social classes use the masculine variant more often, and therefore produce less sharp phonetic distinction between their vowels. For both sexes, vowel distinctiveness appears to correlate with social class, but the correlations are very weak and it is questionable as to whether or not these results can be reproduced in another study.

A comparison of the Lines of Best Fit in the scatterplots in Figure 6.1 shows that Social Class Hypothesis 2 is clearly supported. We predicted that the working-class speakers would show greater sex differentiation compared to the middle-class speakers. If the two Lines of Best Fit for the sexes were overlaid on top of each other, then they would be farthest apart on the left-hand side, among the lower-class speakers, and closest together on the right-hand side, among the middle-class speakers. This signifies that the working-class speakers are making a greater differentiation between the sexes than the middle-class speakers.

6.3.5 Preliminary results for BVD

Again, we will start by comparing the average BVD by sex for each of the dialects. Similar to FVD, we expect that in every case the females should show a greater average BVD than the males. Table 6.3 lists the average BVD in mels by sex for each of the dialect regions. Again, not every region has data for both males and females, and one of the regions, North Central, lacks both. Similar to FVD, the overall trend is for females to

---

12 See footnote 13 for the definitions of outlier and extreme outlier.
have a larger BVD than males, 229.9 mels for the females compared to 212.8 mels for the males, indicating that women produce /uw/ and /o/ with greater phonetic distinctiveness than men. The difference between the two means is statistically significant (t(436) = 4.63, p < 0.001). This is consistent with the overall mean sex differences seen for FVD — for both metrics, women show greater mean distances between vowels than males, and further supports Gender Hypothesis 1.
I tested for a relationship between SEI and BVD with Pearson $r$ correlations after removing the 18 males and 58 females that lack SEI scores. The same results were obtained as for FVD — the ability of a speaker’s SEI to predict their BVD was significant neither for the males ($r(145) = 0.017, p = .841$), nor the females ($r(213) = 0.003, p = .970$). Once again we cannot draw any conclusions without first considering...
the possibility of confounding factors due to ongoing language changes.

6.3.6 Reanalysis of BVD without the confounding regions

Similar to the FVD metric, each of the dialect regions must be considered individually for confounding factors. Most of the dialects show the overall trend of females producing larger mean BVD scores than their corresponding males. However, there are three dialect regions that show the reverse trend: Other Canada, Other North, and Other Western Pennsylvania. In none of these cases are the differences notable.

Similar to the front vowels, the back vowels are also undergoing massive changes in North American English. We need to exclude any dialect regions with back vowels that are undergoing language change. Again, we will turn to the dialect maps in the ANAE to determine which of the regions have stable back vowels.

the relative height of /o/ (p. 86)

There is a tendency for this vowel to be lower in the north and higher in Canada and the West, reflecting the fact that /o/ is raised and backed when merged with /oh/. However, on the whole the height of /o/ is not a distinguishing feature of any dialect area.

the relative fronting and backing of /o/ (p. 87)

The low back vowel is fronting as the second stage of the Northern Cities Shift, but in the northern states only, particularly the Great Lakes area, New York state, and to a lesser extent, Western New England and Midland and Mid-Atlantic.

the relative height of /uw/ (p. 100, 102)

Because the variation seen in /uw/ depends on the preceding environment, two sets of maps are provided for /uw/: one set for /uw/ after coronal consonants, and another set for /uw/ after non-coronal consonants. An examination of the map for /uw/ after coronal consonants shows that one region, the Mid-Atlantic, particularly stands out as having extremely low nuclei. The North and Canada have highest onsets, showing the least amount of differentiation between nucleus and glide.

the relative fronting and backing of /uw/ (p. 101, 103)

The ANAE reports that the fronting of /uw/ after coronal consonants is the most widespread tendency across all North American dialects. In fact, there are only two areas
that do not participate in this sound change: the north central states of Wisconsin and Minnesota, and a north-eastern area extending from western New York thorough New Jersey and into New England. The fronting of /uw/ after non-coronals has not advanced quite as far, but still encompasses all of the South (including the western and eastern regions), Midland, and Mid-Atlantic states. The North, New England and Canada show little or no fronting.

In their description of the variation in /o/ and /uw/ in the dialects of North American English, Labov et al. refer to four sources of variation: the Northern Cities Shift, the Low Back Vowel Merger, the differentiation between nucleus and glide, and the fronting of /uw/. Of these, the Northern Cities Vowel Shift, the Low Back Vowel Merger, and the fronting of /uw/ are ongoing language changes that may confound any results. The following regions are participating in the Northern Cities Vowel Shift, and therefore must be eliminated: North, Western New England, Midland and Mid-Atlantic. The South, South East, and West must also be eliminated because they participate in the fronting of /uw/. This leaves Canada and Eastern New England. Of these, Eastern New England is currently in the final stages of the Low Back Vowel Merger, but some speakers still make the distinction, even in the north New England area (ANAE, Map 16.3). Canada, on the other hand, shows much more consistency, with every Canadian speaker reporting a merger (see §6.4.2 for discussion of the data) in both perception and production before /n/, except the oldest speaker in Vancouver, who reported that they were “close” in perception (ANAE, Map 15.1). The more conservative environment of before voiceless stops show slightly more variation (ANAE, Map 15.2), but even here every speaker but two reports that /o/ and /oh/ either sound the “same” or “close” in production and perception. As such, Canada is the only good candidate with which to proceed with our investigation of the relationship between the BVD index and Socio-Economic Index.

The Canadian regions contain data for thirteen males, of which five are missing SEI scores, and twenty females, of which two are missing SEI scores. The correlation between FVD and SEI for the 26 speakers with SEI scores is significant ($r(24) = .441$, $p = .024$). When we break down the data by sex, then the correlation for the eight males is significant ($r(6) = .765$, $p = .027$), but the correlation is not significant for the eighteen
females \((r(16) = .275, p = .269)\). Similar to FVD, the positive \(r\) values for both of the correlations with BVD indicates that as social class increases, the distance between the two back vowels increases. Figure 6.2 shows the scatterplots for BVD and SEI for the men and the women.

Again, these results give some support to *Social Class Hypothesis 1*, that vowel distinctiveness correlates with social class. Although the results are statistically
significant for the men only, there is a relationship between BVD and SEI for the women as well, as shown by the Lines of Best Fit for the scatterplots in Figure 6.2. So, for both groups, the higher the SEI, the greater the distance between the high and mid back vowels. But again that these results can be reproduced in a another study is questionable.

Unlike the results for front vowel distinctiveness, Social Class Hypothesis 2 is clearly not supported for back vowel distinctiveness. An examination of the Lines of Best Fit in the scatterplots in Figure 6.1 shows that the Lines of Best Fit are more dispersed for the middle-class speakers than the working-class speakers, the opposite of what the hypothesis predicts.

6.3.7 Summary of the results

In general, the females produce vowels that are more distinct than those of the males, as seen in the higher average FVD and BVD metric scores for the females. This adds further support to the claim that women articulate their speech so that the phonetic correlates of the phonological categories are farther apart in acoustic space than the phonetic correlates produced by men.

Socio-Economic Index was used as an approximation of social class. For both the FVD metric and the BVD metric, the males showed a correlation with their SEI scores. In the case of the FVD metric, the correlation was marginally significant, while in the case of the BVD metric, the correlation was significant. The females on the other hand, only showed a significant correlation with FVD, and even then it was marginal. It is difficult to explain why the males show stronger correlations than the females. Given the large number of speakers for the front vowel correlations, and the consistency of the trend across both the front vowels and the back vowels, the sex difference just might be a real difference in the speech of men and women, but why this should be is not clear from the results presented here. It may be that the females are more sensitive to the circumstances of the interview, viz., a telephone interview with a stranger who has identified himself or herself as an academic. Or it may be that phonetic distinctiveness carries greater social significance for males than females, in other words, it is masculine engendered variation only (as defined in §2.7).

The final hypothesis tested was that working-class speakers show greater sex
differentiation for the phonetic distinctiveness measures than middle-class speakers. This hypothesis was not supported. This hypothesis was formulated based on the results of only a few studies, so it may not hold.

The results presented here provide different degrees of support for several of the hypotheses tested, but not all. The most fundamental of the hypotheses, that vowel distinctiveness carries social significance with regard to social class, is somewhat supported. Obtaining the same result for both the FVD metric and the BVD metric adds further weight to our claims, but even still caution is needed as it questionable that this result can be reproduced. In summary, phonetic distinctiveness of vowels in North American English corroborates the theory of phonetic-level engendered variation presented in the first four chapters in so far as the women in general produce more distinct vowels than the men, but the way this sex-based difference interacts with social class clearly needs more work. Further evidence of the relevance of phonetic distinctiveness is presented in the following discussion on vowel mergers.

6.4 Vowel mergers and engendered variation

6.4.1 Hypotheses tested

The other source of data in the ANAE used to test hypotheses presented in the first four chapters is the data on vowel mergers. There are several ongoing vowel mergers in North American English. These changes allow us to test part of the one hypothesis that deals directly with language change:

LV&C Hypothesis 2 (page 118)

Women lead men in linguistic changes that involve the maintenance of phonetic distinctiveness, while men lead women in linguistic change that involves the loss of phonetic distinctiveness, as long as there are not overriding factors (such as prestige) involved.

That women lead sound shifts in general has been well established by Labov and others. In this section, we are concerned with the second half of the hypothesis — men lead women in linguistic change that involves the loss of phonetic distinctiveness. Because vowel mergers involve the loss of the phonetic distinctiveness, they make an ideal testing ground for this hypothesis.
6.4.2 The data

The ANAE investigates six categories of mergers:

- the merger of /a/ and /o/ (COT-CAUGHT) in various environments
- the merger of several tense and lax vowels before [l]
- the merger of /i/ and /ɛ/ before [n] (PIN-PEN)
- the merger of front vowels before intervocalic /r/ (MARY-MARRY-MERRY)
- the merger of /ɔɪ/ and /oɪ/ (HORSE-HOARSE)
- and the merger of /w/ and /ʍ/ (WEAR-WHERE)

The merger data are in the form of impressionistic judgments on minimal pairs. The interviewer first elicited judgements on the perception of a set of minimal pairs from the speaker, who rated each pair as either “the same,” “close to the same,” or “unique.” These responses correspond to the categories “merged,” “close” and “distinct.” The analyst also impressionistically rated each production of minimal pairs by the speaker as either “merged,” “close,” or “distinct.” The following questions are an example of the format used to elicit a minimal pair (in this case, pin and pen):

What would you use to sign a check with?
What would you use to fasten a cloth diaper? A safety …
Do those words sound the same to you?
Say them again for me, and tell me which is which.

Due to time limitations, only one minimal pair was elicited for each contrast tested. The list of environments tested and the words used to elicit the contrast are provided in Table 6.4. For comparison with the dialects maps, etc., provided in the ANAE, the notational system used in the ANAE is listed in the table, followed by my own notation based on actual words. The mergers of the low back vowels are collectively referred to by their familiar name of COT–CAUGHT merger. The COT–CAUGHT merger specifically in the environment of a /t/ is referred to as the NOT–NAUGHT merger.

A broad transcription gloss is also provided for readers unfamiliar with the notational system used in the atlas. The words in the third column come from the one
For each of the mergers listed in Table 6.4, the number of speakers of each sex in each of the three categories “merged,” “close,” and “distinct” was tabulated for both the perception and the production data. The counts for the perception data are provided in Table 6.5. The counts for the production data are provided in Table 6.6. The percentages are the percent of speakers of that sex in that category for that merger, so that each row in the table totals to 100 percent. The percentages allow us to compare directly the men with the women.

For each merger, a Pearson’s Chi square test of significance was carried out to
### Perception Results

<table>
<thead>
<tr>
<th>Merger</th>
<th>Sex</th>
<th>Merged</th>
<th>Close</th>
<th>Distinct</th>
<th>Test of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DON–DAWN</td>
<td>female</td>
<td>50.2</td>
<td>9.8</td>
<td>40.0</td>
<td>7.96, 0.019</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>44.2</td>
<td>16.9</td>
<td>45.0</td>
<td></td>
</tr>
<tr>
<td>DOLL–TALL</td>
<td>female</td>
<td>52.1</td>
<td>6.9</td>
<td>41.0</td>
<td>15.32, 0.001</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>45.2</td>
<td>17.0</td>
<td>43.0</td>
<td></td>
</tr>
<tr>
<td>SOD–SAWED</td>
<td>female</td>
<td>42.3</td>
<td>10.1</td>
<td>47.6</td>
<td>3.33, 0.189</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>48.0</td>
<td>15.0</td>
<td>37.0</td>
<td></td>
</tr>
<tr>
<td>COT–CAUGHT</td>
<td>female</td>
<td>43.4</td>
<td>8.5</td>
<td>48.2</td>
<td>8.68, 0.013</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>36.6</td>
<td>15.0</td>
<td>48.4</td>
<td></td>
</tr>
<tr>
<td>HOCK–HAWK</td>
<td>female</td>
<td>33.1</td>
<td>10.2</td>
<td>56.7</td>
<td>0.41, 0.815</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>35.5</td>
<td>9.4</td>
<td>55.1</td>
<td></td>
</tr>
<tr>
<td>HILL–HEEL</td>
<td>female</td>
<td>10.5</td>
<td>7.3</td>
<td>82.2</td>
<td>0.23, 0.891</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>11.7</td>
<td>6.8</td>
<td>81.5</td>
<td></td>
</tr>
<tr>
<td>FULL–FOOL</td>
<td>female</td>
<td>12.0</td>
<td>7.5</td>
<td>80.5</td>
<td>9.63, 0.008</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>12.4</td>
<td>14.6</td>
<td>73.0</td>
<td></td>
</tr>
<tr>
<td>BELL–BAIL</td>
<td>female</td>
<td>10.8</td>
<td>7.6</td>
<td>81.6</td>
<td>1.37, 0.504</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>10.9</td>
<td>10.4</td>
<td>78.7</td>
<td></td>
</tr>
<tr>
<td>PIN–PEN</td>
<td>female</td>
<td>28.7</td>
<td>10.9</td>
<td>60.4</td>
<td>6.66, 0.036</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>36.2</td>
<td>13.2</td>
<td>50.6</td>
<td></td>
</tr>
<tr>
<td>MERRY–MARY</td>
<td>female</td>
<td>61.5</td>
<td>4.1</td>
<td>34.4</td>
<td>12.04, 0.002</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>69.1</td>
<td>10.1</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>MERRY–MARRY</td>
<td>female</td>
<td>81.0</td>
<td>8.1</td>
<td>10.9</td>
<td>0.34, 0.844</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>83.3</td>
<td>7.3</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>HOARSE–HORSE</td>
<td>female</td>
<td>1.5</td>
<td>2.0</td>
<td>96.5</td>
<td>2.67, 0.263</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>3.5</td>
<td>3.5</td>
<td>94.8</td>
<td></td>
</tr>
<tr>
<td>WITCH–WHICH</td>
<td>female</td>
<td>85.6</td>
<td>2.0</td>
<td>12.4</td>
<td>2.70, 0.259</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>84.2</td>
<td>3.5</td>
<td>12.3</td>
<td></td>
</tr>
</tbody>
</table>

Note: d.f. = 2 for all Chi square tests

Table 6.5 Counts and row percentages for three degrees of speaker perception of a merger, “merged,” “close” and “distinct” for men and women for thirteen ongoing mergers in North American English. For each merger, a Pearson Chi Square test of significance was carried out on the two sexes. Chi square and p values are provided for each of the mergers.
### Production Results

<table>
<thead>
<tr>
<th>Merger</th>
<th>Sex</th>
<th>Merged</th>
<th>Close</th>
<th>Distinct</th>
<th>Test of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>DON–DAWN</td>
<td>female</td>
<td>46.6</td>
<td>211</td>
<td>11.9</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>42.0</td>
<td>113</td>
<td>12.6</td>
<td>34</td>
</tr>
<tr>
<td>DOLL–TALL</td>
<td>female</td>
<td>35.8</td>
<td>141</td>
<td>22.1</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>37.2</td>
<td>81</td>
<td>18.8</td>
<td>41</td>
</tr>
<tr>
<td>SOD–SAWED</td>
<td>female</td>
<td>40.5</td>
<td>68</td>
<td>14.9</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>46.0</td>
<td>46</td>
<td>18.0</td>
<td>18</td>
</tr>
<tr>
<td>NOT–NAUGHT</td>
<td>female</td>
<td>33.1</td>
<td>154</td>
<td>15.1</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>34.2</td>
<td>94</td>
<td>12.0</td>
<td>33</td>
</tr>
<tr>
<td>HOCK–HAWK</td>
<td>female</td>
<td>31.8</td>
<td>122</td>
<td>11.5</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>27.4</td>
<td>64</td>
<td>12.8</td>
<td>30</td>
</tr>
<tr>
<td>HILL–HEEL</td>
<td>female</td>
<td>7.2</td>
<td>31</td>
<td>8.6</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>10.0</td>
<td>25</td>
<td>6.0</td>
<td>15</td>
</tr>
<tr>
<td>FULL–FOOL</td>
<td>female</td>
<td>7.7</td>
<td>35</td>
<td>10.4</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>9.0</td>
<td>24</td>
<td>10.8</td>
<td>29</td>
</tr>
<tr>
<td>BELL–BAIL</td>
<td>female</td>
<td>6.3</td>
<td>24</td>
<td>8.2</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>5.7</td>
<td>12</td>
<td>9.4</td>
<td>20</td>
</tr>
<tr>
<td>PIN–PEN</td>
<td>female</td>
<td>25.7</td>
<td>116</td>
<td>13.7</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>28.3</td>
<td>75</td>
<td>15.5</td>
<td>41</td>
</tr>
<tr>
<td>MERRY–MARY</td>
<td>female</td>
<td>64.4</td>
<td>159</td>
<td>6.9</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>73.8</td>
<td>110</td>
<td>4.0</td>
<td>6</td>
</tr>
<tr>
<td>MERRY–MARRY</td>
<td>female</td>
<td>86.3</td>
<td>214</td>
<td>5.2</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>86.0</td>
<td>129</td>
<td>5.3</td>
<td>8</td>
</tr>
<tr>
<td>HOARSE–HORSE</td>
<td>female</td>
<td>1.3</td>
<td>5</td>
<td>1.0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>0.4</td>
<td>1</td>
<td>1.7</td>
<td>4</td>
</tr>
<tr>
<td>WITCH–WHICH</td>
<td>female</td>
<td>83.7</td>
<td>343</td>
<td>2.0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>82.3</td>
<td>191</td>
<td>4.1</td>
<td>11</td>
</tr>
</tbody>
</table>

Note: d.f. = 2 for all Chi square tests

Table 6.6 Counts and row percentages for three degrees of speaker production of a merger, “merged,” “close” and “distinct” for men and women for thirteen ongoing mergers in North American English. For each merger, a Pearson Chi Square test of significance was carried out on the two sexes. Chi square and p values are provided for each of the mergers.
determine if the men and the women are behaving significantly differently from each other. The last two columns in Table 6.5 and Table 6.6 contain the values for $\chi^2$ and $p$ for the mergers that show significant sex differences. Altogether, six of the perception mergers show significant sex differences, but none of the production mergers show significant sex differences. In general, the females produce more merged and close minimal pairs for the COT–CAUGHT distinctions, while males produce more merged and close minimal pairs for the /in/-/en/, the tense-lax distinction before /l/, and the distinction of front vowels before intervocalic /r/. The data for the HOARSE–HORSE distinction and the WITCH–WHICH distinction are ambiguous as to which sex is leading the change. The only clear exception to this pattern is for SOD–SAWED: men lead this merger even though women lead all of the other COT–CAUGHT mergers, although the differences are not significant and therefore may be a statistical anomaly. The sex differences are discussed in more detail in the following section.

Before looking at the sex differences in more detail, we must first address the differences between perception and production. The perception data show significant differences between the sexes for five of the mergers. The production data, however, do not contain a single merger that shows a significant sex difference. A comparison of the percentages in the merged column of the perception results with the percentages in the merged column of the production results reveals that more mergers are perceived than actually produced with the exception of the mergers before intervocalic /r/. This implies that two phonemes are perceived as being merged before they are produced as merged. Labov et al. (2006:62) comment that:

The study of sound changes in progress shows that the relations of production and perception are not in general symmetrical. In the majority of cases, the change occurs earlier in perception than in production…

It seems that sex differences in mergers initially appear in perception, but the trailing sex quickly makes up the difference in production. This suggests that prestige is not involved in the merger; if one of the variants was associated with social prestige, then we would expect the sex differences to increase in production, as women adapted the prestigious variant more than the men. We must look to some other explanation other than prestige to account for the sex differences.
6.4.4 Testing the hypothesis

The perception mergers that show a significant sex difference reveal that men and women are leading an equal number of mergers. This is summarized in Table 6.7. Although only these six mergers show significant sex differences, they are emblematic of the general pattern: women lead the COT–CAUGHT mergers, while men lead the PIN–PEN merger, the mergers of tense and lax vowels before /l/, and the merger of front vowels before intervocalic /r/.

The hypothesis that we are testing is that men lead women in linguistic change that involves the loss of a phonetic-level distinctions, as long as there are no overriding factors (such as prestige) involved. There are two components of the hypothesis that need to hold before we can claim that the hypothesis is valid. The first component is that these changes do not involve overriding factors, such as the avoidance of stigmatized forms. The second component is that for all of the mergers for which the first point holds true, men lead mergers more often than women. We will begin with the first point.

Conveniently for us, the authors of the ANAE considered whether or not the mergers were taking place below the level of social awareness. For each region, they compare the percentage of speakers who report the minimal pairs as the “same” for each region with the percentage of speakers who produce the minimal pairs as merged. To illustrate their point they provide two graphs, one showing the percentages of speakers with fully merged COT–CAUGHT in perception by region, and the other showing the percentages of speakers with fully merged COT–CAUGHT in production by region. The percentages are shown Figure 6.3. A comparison of the perception data with the production data shows that the lines connecting the percentages overlap. Labov et al. comment on the similarity between the production and the perception data as follows (ANAE, p.62):

The results are almost identical. This indicates that the speaker’s judgments are not heavily affected by any tendency towards correction of a stigmatized form or exaggeration of a prestige norm. Like other mergers, the fusion of /o/ and /oh/ takes place below the level of social awareness and is normally not the focus of sociolinguistic evaluation.

The assumption here is that a notable difference between the percentage of speakers who report the minimal pairs as merged and the percentage of speakers who produce the
minimal pairs as merged would be due to overcorrection on behalf of the speakers. This is only possible if the speakers are socially aware of the merger. Given this, we can assume that the mergers are taking place below the level of social awareness, satisfying the first requirement.

The second requirement, that men are leading the mergers, is not so straightforward. According to Table 6.7, it is clearly not the case that men are leading all of the mergers — women are leading the merger of COT–CAUGHT. This seems to refute LV&C Hypothesis 2. However, a closer examination of the interaction between the mergers and the rest of the vowel system reveals that the picture is much more complicated than initially presumed. Consider the following two quotations from the ANAE describing the interaction between the rest of the vowel system and the merger of COT–CAUGHT:

One of the major events in the differentiation of North American dialects is the low back merger of /o/ and /oh/. In some areas, particularly Canada, this event triggers a chain shift among the front short vowels, which have been relatively stable over long periods of English history. (ANAE, p.19)

The merger of /o/ and /oh/ affects all dialects except those where it is blocked by some specific result of the sound changes in progress. In the Inland North it is blocked by the fronting of /o/; in New York City and the Mid-Atlantic dialect, by the raising of /oh/; and in the South, by the development of the back upglide which forms part of the /aw/ shift. (ANAE, p.147)

The other mergers that show significant sex differences do not show interactions with the vowel system — they neither initiate changes, nor are they blocked by changes. So it seems that the COT–CAUGHT merger is involved in the maintenance of phonetic-level distinctions indirectly, in the sense that the merger either alters or is blocked by other

<table>
<thead>
<tr>
<th>Merger</th>
<th>Female-led</th>
<th>Male-led</th>
</tr>
</thead>
<tbody>
<tr>
<td>DON–DAWN</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DOLL–TALL</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>NOT–NAUGHT</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>FULL–FOOL</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>PIN–PEN</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>MERRY–MARY</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.7 The six perception mergers that show significant sex differences. Women are leading three of the mergers, while men are leading the other three.

The second requirement, that men are leading the mergers, is not so straightforward. According to Table 6.7, it is clearly not the case that men are leading all of the mergers — women are leading the merger of COT–CAUGHT. This seems to refute LV&C Hypothesis 2. However, a closer examination of the interaction between the mergers and the rest of the vowel system reveals that the picture is much more complicated than initially presumed. Consider the following two quotations from the ANAE describing the interaction between the rest of the vowel system and the merger of COT–CAUGHT:

One of the major events in the differentiation of North American dialects is the low back merger of /o/ and /oh/. In some areas, particularly Canada, this event triggers a chain shift among the front short vowels, which have been relatively stable over long periods of English history. (ANAE, p.19)

The merger of /o/ and /oh/ affects all dialects except those where it is blocked by some specific result of the sound changes in progress. In the Inland North it is blocked by the fronting of /o/; in New York City and the Mid-Atlantic dialect, by the raising of /oh/; and in the South, by the development of the back upglide which forms part of the /aw/ shift. (ANAE, p.147)

The other mergers that show significant sex differences do not show interactions with the vowel system — they neither initiate changes, nor are they blocked by changes. So it seems that the COT–CAUGHT merger is involved in the maintenance of phonetic-level distinctions indirectly, in the sense that the merger either alters or is blocked by other
The COT–CAUGHT distinction is an intrinsic part of the English vowel system, and the merger is closely connected to other distinctions in the vowel system. In comparison, the other mergers all take place in isolation. This is reminiscent of Labov’s observation that the language changes that men lead are isolated shifts (see quotation from Labov on page 124).

We can conclude from this that the original wording of LV&C Hypothesis 2 was too strong. As it was originally worded, it claimed that men will always lead sound changes that result in the loss of phonetic-level distinctions. However, the data presented here show that this is not the case. We must change the wording:
The new wording of the hypothesis now includes the possibility that the sound changes that women lead will be indirectly linked to the maintenance of phonetic-level distinctions. This wording encompasses mergers such as COT–CAUGHT, since it is indirectly linked to the maintenance of phonetic-level distinctions, but the new wording does not encompass the other mergers since they are not linked to the maintenance of phonetic-level distinctions, either directly nor indirectly.

6.5 Linking vowel distinctiveness and degree of merger

6.5.1 Initial considerations

In the previous two sections, I argued that men produce less phonetically-distinct vowels than women, and that men lead vowel changes that result in the loss of phonetic distinctiveness, such as a vowel merger. This implies that there is a link between phonetic distinctness of the vowel system in general, and the degree of a specific merger. If that is the case, then we should see a relationship between the two measures of vowel distinctiveness and the merger categories “merged,” “close” and “distinct.” Specifically, the speakers with smaller FVD and BVD index scores should also be the speakers whose productions of the target words are perceived as merged the most often. In this section I test for such a link. As well, given Lindblom’s (Liljencrants and Lindblom 1972) assumption that a vowel’s influence on neighbouring vowels decreases with distance (see discussion in §6.3.1), then the front vowel mergers such as the PIN–PEN merger should be more sensitive to front vowel distinctiveness than to back vowel distinctiveness. Likewise, the back vowel mergers such as FULL–FOOL should be more sensitive to back vowel distinctiveness than to front vowel distinctiveness.

We can test for a link between vowel distinctiveness and merger categories by
looking for correlations between merger category and the distinctiveness metrics. There are, however, several considerations. The first consideration is the relationship between the merger categories “merged,” “close” and “distinct.” The second consideration is which of the mergers to include. The third consideration is which regions to include.

The relationship between the merger categories “merged,” “close” and “distinct” is ordinal but not linear. The linear regression model does not handle non-linear ordinal variables well. The statistical test for relationships between an ordinal variable as the dependent variable and one or more scalar independent variables is Ordinal Regression Analysis. This analysis is carried out by first transforming the continuous independent variables into discrete categories, and then using the Pearson’s chi square test statistic. The model does not predict the cumulative probabilities directly, but rather a function of those values. The mathematical transformation that then produces the probabilities is called the “link function,” and is chosen to match the distribution of the data among the ordinal categories. The link functions used in this chapter are listed in Table 6.8.

The next consideration is which of the mergers to include. The consonantal merger WITCH–WHICH is excluded as it is a merger between a voiceless phoneme and its voiced counterpart, and therefore has no connection with vowel distinctiveness. As well, in the last section I argued that the COT–CAUGHT mergers are part of other ongoing changes in the vowels. As such, even if we found a correlation between BVD and the degree of merger for these categories, it still would not be clear what this meant. It may be that vowel distinctiveness is related to degree of merger for the COT–CAUGHT mergers, or it may be that the vowels used in the calculation of the BVD index are participating in the vowel changes that are related to the COT–CAUGHT merger, and therefore we are attempting to correlate two measures of the same sound change. The COT–CAUGHT mergers are still included for comparative purposes, but we must be cautious when interpreting their results. The remaining seven mergers (the bottom eight mergers in Tables 6.5 and 6.6 minus the WITCH–WHICH merger) should interact with vowel distinctiveness in the ways predicted.

The final consideration concerns which regions to include for each merger. In §6.3.4 and §6.3.6, the regions with ongoing sound changes that involved the vowels used in the calculation of the FVD and BVD index scores were excluded. These regions are
included here because if there really is a relationship between general vowel distinctiveness and specific vowel mergers, then even if the vowels used in the calculation of the FVD and BVD are undergoing changes, the correlations should still hold. Consider the specific example of the eight speakers from the St. Louis region. These speakers show both the Northern Cities Shift and the PIN–PEN merger. This merger is most common in the south, and therefore is not related to the Northern Cities Shift directly. The Northern Cities Shift alters the position of the low front vowel /æ/, which is used in the FVD calculation. As this vowel moves up towards /i/, the /æ/ vowel applies pressure on the vowels above it, the front vowel space is compressed, and there is a reduction of phonetic distinctiveness of the front vowels. One consequence of this reduction of the front vowel space is that /ɛ/ moves down or back. If vowel mergers are related to vowel distinctiveness, then another consequence may be that /ɛ/ and /ɪ/ begin to merge. If so, then those speakers who have a reduced distance between /æ/ and /i/ due to the shifted position of /æ/ should also be the speakers who show the most advance stages of the PIN–PEN merger.

However, not all regions were included. Regions that show no or limited variation between speakers are problematic. In these regions, the merger is either in its preliminary stages, or it is in its final stages. Because these regions do not show much variation between the speakers regardless of the vowel distinctiveness of the speaker, including these regions runs the risk of concealing the relationship between vowel distinctiveness and merger category. In order to determine which regions should be excluded, each region was assigned an index score representing the degree of merger in that region. The

<table>
<thead>
<tr>
<th>Link Function</th>
<th>Typical Application</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit</td>
<td>each category is roughly equal</td>
<td>log(γ/(1 - γ))</td>
</tr>
<tr>
<td>Complementary log-log</td>
<td>higher categories are more probable</td>
<td>log(-log(1 - γ))</td>
</tr>
<tr>
<td>Negative log-log</td>
<td>lower categories are more probable</td>
<td>-log(-log(γ))</td>
</tr>
<tr>
<td>Cauchit</td>
<td>extreme values are more probable</td>
<td>tan(π(γ - 0.5))</td>
</tr>
</tbody>
</table>

Table 6.8 Various link functions for the Ordinal Regression Analysis
index scores were calculated as follows. Each merger was cross-tabulated for merger category ("merged," “close” or “distinct”) by region. The cross tabulation charts of the the mergers are provided in Appendix D. The index score for a region was calculated by assigning the responses in the “merged” category a weight of two, the responses in the “close” category a weight of one and the responses in the “distinct” category a weight of zero. These weights were then summed and divided by the total number of responses. This resulted in a number between zero and two, which was then multiplied by 100 for ease of presentation. The calculation for the index as a formula is:

$$\text{index} = \frac{( \# \text{of MERGE} \times 2 + \# \text{of CLOSE} ) 	imes 100}{N}$$

A value of 200 represents a complete merger in that region, whereas a value of zero represents a complete distinction.

The index scores were then used to decide which regions to include, with regions showing no or limited variation among the speakers excluded. A region was considered to show no or limited variation if its index score was between 0 and 24, or between 176 and 200. Regions with index scores between 25 and 175 were included in the calculations.

6.5.2 Results

For each of the mergers, two Ordinal Regression Analyses were run, one with FVD as the predictor of merger category, and one with BVD as the predictor of merger category. The merger categories were coded so that “merged” was the lowest and “distinct” was the highest. With this coding scheme a positive correlation implies that as vowel distinctiveness decreases, the vowels involved in the mergers merge more frequently. For each merger, the Link Function was set according to the distribution of the merger categories. The same Link Function was used for both the FVD and the BVD regression analysis, since the distribution of the merger categories does not change between these two.

Table 6.9 lists the results for each of the mergers. For each merger, the following information is provided: the link function used in the statistical test; the number of data items; and for each of FVD and BVD, an estimate of the coefficient; the Chi square statistic; and the significance level of the correlation. Altogether, 16 of the 24
correlations between the vowel distinctiveness indices and the merger categories are marginally significant or significant. Every merger showed at least one significant or marginally significant correlation with a vowel distinctiveness metric except the FULL–FOOL merger, and even this merger participates in the salient patterns discussed in the next paragraph. The implication of this is that vowel mergers do not happen in isolation. However, the interpretation of Table 6.9 is not straightforward.

There are two salient patterns in Table 6.9. The first pattern concerns the strength of the Chi square statistic. First consider the COT–CAUGHT mergers. Setting aside the significance levels for the moment and only considering the size of the Chi square statistic, we see that for every one of the five COT–CAUGHT mergers the correlation with FVD is stronger than BVD. This intimate relationship between front vowel distinctiveness and the COT–CAUGHT merger is further evidence that the COT–CAUGHT merger is closely connected to the position of the front vowels, suggesting that it participates in a vowel shift that involves front vowels as well as back vowels. This further motivates the special treatment of the COT–CAUGHT mergers in §6.4.4.
Of the remaining seven mergers listed in Table 6.9, five involve the front vowels, while the other two involve the back vowels. For the five mergers involving front vowels, the correlation with FVD is stronger than the correlation with BVD. The reverse pattern holds for the two mergers involving back vowels — the correlation with BVD is stronger than the correlation with FVD. The consistency of the pattern across all seven mergers is remarkable. It is also what Lindblom and his collaborators predicted. In their model of vowel space briefly introduced in §6.3.1, Lindblom claimed that a vowel’s influence on other vowels decreases with distance. The patterning of the relationship between the mergers and vowel distinctiveness empirically confirms this. Front vowel mergers are related to the distinctiveness of the front vowels but are not related to (or weakly related to) the distinctiveness of back vowels, while back vowel mergers are related to the distinctiveness of the back vowels but are not related to (or are weakly related to) the distinctiveness of the front vowels.

The second pattern concerns the coefficients. It is difficult to directly interpret the coefficients due to the nature of the link function, but they are useful in that they tell us the direction of the relationship between the merger categories and vowel distinctiveness indices. A positive coefficient indicates a positive relationship, viz., as vowel distinctiveness increases, the merger category increases (recall that the order of the categories from lowest to highest is “merged,” “close,” “distinct”), while a negative coefficient indicates a negative relationship between vowel distinctiveness and merger category. An examination of the signs of the coefficients reveals three trends. The first trend is that the correlations with the COT–CAUGHT mergers all have negative coefficients for both FVD and BVD. This implies that an increase in vowel distinctiveness for both the front vowels and the back vowels corresponds to merging the COT–CAUGHT vowels. Most likely the merging of the two low back vowels results in a slight depopulation of the otherwise crowded English vowel space, allowing all vowels to spread out further. The effect will only be seen for unconditional mergers such as COT–CAUGHT, and not for any of the other mergers, all of which are conditioned by environment.

The second trend in the signs of the coefficients is that the vowel distinctiveness metrics positively correlate with merger category for the mergers before /l/ and /n/. Again,
the consistency of the pattern across both vowel distinctiveness metrics and all four vowel mergers is remarkable. These mergers behave exactly as predicted. They show positive correlations with the vowel distinctiveness metrics, and they are more sensitive to vowel distinctiveness metric in the same region as the merger. The positive correlation with vowel distinctiveness for these four mergers was expected because it implies that the speakers with smaller vowel spaces, and therefore less distinct vowels, are also the speakers who show the merger first. Furthermore, the relationship between vowel distinctiveness and merger categories is significant for three of four of these mergers, and the fourth merger follows all the same trends. This relationship between these mergers and vowel distinctiveness strongly supports the hypothesis that there is a link between phonetic distinctness of the vowel system in general and the degree of a specific merger.

The third trend involves the set of mergers before /r/. Again we see remarkable consistency: all three mergers show the same pattern for both vowel distinctiveness metrics. However, the relationship between vowel distinctiveness and the merger categories for the mergers is in the wrong direction (all of the coefficients are negative). While we expected that the speakers with smaller vowel spaces would also be the speakers who showed the mergers before /r/ first, it is actually the speakers with the larger vowel spaces who show the mergers first. Further complicating this picture is the fact that males are clearly leading the MERRY–MARY merger, as shown in Table 6.5.13 This is regardless of the fact that males have reduced vowel distinctiveness. Clearly mergers before /r/ are more complex than initially assumed. Besides their unexpected relationship with vowel distinctiveness, as I pointed out in §6.4.3 the mergers before /r/ also do not show the “perception leads production pattern” that all of the other mergers show. More detailed sociolinguistic and phonetic studies of the these mergers and their interaction with vowel space dispersion is called for before we can explain these results. There are two specific concerns. Why does production of the /r/ mergers lead perception? Are the mergers taking place above the level of social awareness? If so, then the negative correlation with vowel distinctiveness may be confounded by social class.14 As well,
Labov has noted that the phonological vowel space before /r/ is contracted in comparison to full vowel space of a speaker (Labov 1994:243). What is the relationship between the phonological vowel space before /r/ and the non-contracted phonological vowel space? These questions, and others, need to be addressed before an adequate account of mergers before /r/ can be presented.

6.6 Chapter summary

This chapter introduces data from the Atlas of North American English with the objective of testing several of the engendered variation framework hypotheses. The first part of the chapter used the concept of vowel distinctiveness to test Gender Hypothesis 1 introduced at the end of Chapter Three. The second part of the chapter tested the LV&C Hypothesis 2 against the vowel mergers data. The last part of the chapter tested for a direct link between the general vowel distinctiveness of a speaker and specific vowel mergers.

In general, women as a group have greater average vowel distinctiveness scores than men for both the front vowels and the back vowels. This supports Gender Hypothesis 1, that women will articulate their speech so the phonetic correlates of the phonological categories will be farther apart in acoustic and/or temporal space than the phonetic correlates produced by men.

Social Class Hypothesis 1 claims that vowel distinctiveness correlates with social class. This is based on the claim that working-class speech is more masculine than middle-class speech. After the potentially confounding regions were eliminated, the FVD scores showed a marginally significant correlation with social class for both men and women. The BVD scores showed a significant correlation with social class for men but not for women. The slope of the Line of Best Fit was positive for all four correlations, indicating that as social class increased vowel distinctiveness increased. These results support Social Class Hypothesis 1, and taken together with the results of the last chapter, we can conclude that phonetic-level distinctiveness is used by speakers to index social constructs such as social gender and social class.

The second social class hypothesis claims that working-class speakers show greater sex differentiation for both masculine and feminine engendered variants than middle-class speakers. We expected to see greater inter-group differences between the sexes for
the vowel distinctiveness metrics of the working-class speakers than the middle-class speakers. The predicted pattern was seen for front vowel distinctiveness, but the reverse pattern was seen for back vowel distinctiveness. At this point, it is unclear why we failed to find clear supporting evidence for this hypothesis. It may be that this hypothesis does not hold for phonetic-level engendered variation because it takes place below the level of social awareness, or it may be that the lower class women were reducing the social distance between themselves and the interviewer by speaking in a more distinct manner, but the lower class men were not doing this.

The last hypothesis tested is that women lead men in linguistic changes from below that involves the maintenance of a phonetic-level distinctiveness, while men lead women in linguistic change that involves the loss of a phonetic-level distinctiveness. This was tested by examining several vowel mergers in North American English. Data was provided for both the perception of the mergers and the production of the mergers. With the exception of the mergers before /r/, the percentage of speakers merged was consistently higher for the perception data in comparison to the corresponding production data, showing that perception of a merger leads production of a merger. Furthermore, significant sex differences were only seen in perception data, suggesting that there is an initial sex difference in the perception, but the trailing sex quickly makes up the difference in production. This reduction of sex differences in production suggests that prestige is not a driving force behind the mergers. This dovetails neatly with my claim that the phonetic distinctiveness is the driving force behind sound changes as long as they are void of prestige.

Six of the perception mergers showed significant sex differences, while none of the production mergers showed significant sex differences. Of the mergers that show significant sex differences, women lead the three COT–CAUGHT mergers, while men lead the rest. These results, along with the fact that the COT–CAUGHT merger is indirectly connected to other changes in the vowel system while the other mergers are not, motivated the revision of the hypothesis to the following: women lead men in linguistic changes from below that are linked to the maintenance of a phonetic-level distinctiveness either directly or indirectly, while men will lead women in linguistic change that involves the loss of a phonetic-level distinctiveness.
Lastly, a direct link between the general phonetic distinctiveness of vowels and specific vowel mergers was established. Vowel mergers are sensitive to the dispersion of the vowels in their vicinity, so that for example, the PIN–PEN merger correlates with the distance between /ɪ/ and /æ/. This link establishes the importance of the role of phonetic distinctiveness in vowel mergers.

Collectively, the results presented here corroborate the engendered phonetic-level variation framework, at least for the behavioural sex differences patterns introduced in Chapter Three. Furthermore, these results deepen our understanding of language variation and change in North American English:

- women articulate vowels such that their phonetic distinctiveness is greater than the vowels in the speech of men
- the degree of merger is closely associated with the phonetic distinctiveness of vowels in the surrounding vicinity of the vowel space
- this sex difference in phonetic distinctiveness has social significance
- the sex difference in phonetic distinctiveness helps account for why women tend to lead sound changes that are below the level of social awareness
- the sex difference in phonetic distinctiveness accounts for some of the exceptions to Labov’s principle of that women lead sound changes below the level of social awareness; specifically, it explains why men lead certain mergers.

The last point is particularly important, as it gives further explanatory power to Labov’s principle by accounting for the exceptions.
Conclusions

Everyone agrees that gender is a social factor — language is not differentiated by the biological aspects of sex differences.

Labov 2001a:26

7.1 Revisiting the engendered–nongendered distinction

7.1.1 The engendered–nongendered distinction

The distinction between engendered and nongendered phonetic-level variation as presented in Figure 1.1 forms the backbone of this thesis. If a variable is anatomically-determined engendered variation, then the sex differentiation of the variation is largely predictable. This holds regardless of speech community, since sex differences in the vocal tract are universal — the anatomies of the vocal tracts of men and women in Japan differ in the same way as the vocal tracts of men and women in Canada. Therefore the anatomically-determined sex-based patterns are also consistent across dialects and languages.

I introduced several hypotheses based on the engendered–nongendered distinction. This set of hypotheses is not intended to be complete. Rather than producing an exhaustive theory of engendered phonetic-level variation, my objective was to produce enough evidence to justify the engendered–nongendered distinction. I have captured the essence of the role these sex differences play in phonetic-level variation in these hypotheses. I hope that this set of hypotheses is sufficient to make researchers realize the significance of the phonetic-level engendered variation framework.

For the same reason that I do not present a complete set of hypotheses on phonetic-level engendered variation, I do not directly test all of the hypotheses introduced in this dissertation. Instead, I focus on a few key hypotheses. Four of them concern phonetic
distinctiveness, and lead to greater insight into the role of sex and gender in language variation and change. They are discussed in more detail in the following two sections.

7.1.2 Phonetic distinctiveness as a sociolinguistic variable

In Chapter Two, I laid out and exemplified the characteristics of anatomically-determined phonetic-level engendered variation with a review of five phonetic-level engendered phenomena: VOT, sibilant centre of gravity, creak, breathiness, and pitch. The chapter concluded with a list of the characteristics of anatomically-determined phonetically-level engendered variation. The two most important are the cross-linguistic consistency of the sex-based patterns, and the frame for the indexation of social gender created by these sex-based patterns. The first of these points is based on the review of a limited number of studies of languages other than English. The second of these points has been empirically demonstrated in the past, and was further corroborated in Chapter Five.

I continued to build on the engendered–nongendered distinction in Chapter Three with an examination of cross-linguistic sex-based patterns that are not grounded on anatomical differences. The chapter began with the observation that women speak more clearly than men. This equates to women making sharper phonetic distinctions than men, where “phonetic distinctiveness” refers to the acoustic or temporal distinction between the phonetic realizations of two related phonological constituents, such as a phonological contrast. I illustrated the greater phonetic distinctiveness of women with examples from overall vowel space dispersion, vowel and consonantal reduction, phonemic vowel length, and allophonic variation. In every case examined, women produced sharper phonetic distinctions than men for the same phonemic or allophonic contrast. Furthermore, the studies were from a variety of unrelated languages such as Japanese, Arabic, and Creek, suggesting that there is cross-linguistic consistency in the sex-based patterns of phonetic distinctiveness. These observations led to the proposition of the following hypothesis on phonetic distinctiveness:

Gender Hypothesis 1 (page 84)

The phonetic distinctiveness of the speech of women is greater than that of the speech of men.

This is the first of the four central hypotheses on phonetic distinctiveness.
My intention is that phonetic distinctiveness is one type of behaviourally-determined engendered variation, and that it has all of the characteristics of engendered variation. This includes the claim that sex-based differences in our speech form a frame for the indexation of social gender. As I mentioned in §2.3.1, this point has been demonstrated empirically for at least one of the anatomically-determined sex-based patterns, that men produce lower frequency sibilants than women (Stuart-Smith et al. 2003), but not for a behaviourally-determined sex-based pattern such as phonetic distinctiveness. This was the second hypothesis on phonetic distinctiveness:

**Gender Hypothesis 2 (page 86)**

Phonetic distinctiveness indexes social gender.

I see *Gender Hypothesis 2* as being one of the fundamental ideas presented in this dissertation.

*Gender Hypothesis 2* formed the impetus behind the empirical investigation undertaken in Chapter Five. In that chapter, I examined the temporal distinctiveness of contextually long and short vowels of eight DJs. For each DJ, I measured the ratio of short vowel duration to long vowel duration under two stress conditions, i.e., how distinct the two phonological categories were from each other in the temporal domain. The ratios of the contextual vowel lengths correlated with naïve ratings of the perceived machismo of the DJs, corroborating *Gender Hypothesis 2*.

The implication of this result is that there is stable phonetic-level variation in our speech. While there are relationships between phonological constituents in our speech that are undergoing change, most relationships are not undergoing change. Yet, all phonological constituents, regardless of whether or not they are participating in a sound change, participate in paradigmatic and syntagmatic relationships with other phonological constituents. Therefore phonetic distinctiveness is a ubiquitous component of our speech. One of Labov’s principles is concerned with stable variation: “In stable sociolinguistic stratification, men use a higher frequency of non-standard forms than women” (Labov 1990:205). This principle is concerned with variables that have a prestige component (see discussion in §4.3.1), and therefore does not apply to phonetic-level engendered variation in which prestige is not a factor.
7.1.3 Phonetic distinctiveness and language change

Chapter Four illustrates the significance the engendered–nongendered distinction variation has for sociolinguistic theory. Once we acknowledge that at the phonetic level behavioural and anatomical differences between men and women play a role in determining how we speak, then we see that these differences also play a role in phonetic-level language variation in predictable and socially significant ways. As I mentioned in the last section, this point is particularly important for language variation that is taking place below the level of social significance. Engendered variation that is void of social significance (i.e., none of the variants is considered more prestigious than the others) still has associations with masculinity and / or femininity. The engendered component along with the lack of prestige component results in an increase in predictability of the patterns. For example, because working-class speech is associated with masculinity, we predict that working-class speech will contain more of masculine variants than middle-class speech. If there is a prestige component to the variation, then we must determine which variant is prestigious, and to whom, both highly specific to the speech community. Without the prestige component, we are only left with the engendered component. Because the phonetically-determined sex-based patterns are consistent across languages, dialects, and speech communities, the way phonetic-level engendered variation patterns should also be consistent across languages, dialects, and speech communities. These observations led to the proposition of the following pair of hypotheses on engendered variation and language variation.

LV&C Hypothesis 1a (page 109)
Men will use a masculine variant more frequently than women regardless of whether the variation is stable sociolinguistic stratification or a change-in-progress, as long as it is void of prestige.

LV&C Hypothesis 1b (page 109)
Women will use a feminine variant more frequently than men regardless of whether the variation is stable sociolinguistic stratification or a change-in-progress, as long as it is void of prestige.

In the last section of Chapter Four, I introduced three of Labov’s principles on the role of sex in language variation and change. I argue that one of these principles, namely that women lead linguistic change that occur below the level of social consciousness,
does not necessarily hold at the phonetic level with respect to engendered variation. Even though the variation is below the level of social awareness and therefore void of prestige, I assume that it still has the engendered component. For example, the voicing of /h/ will still be associated with masculinity, regardless of whether it is involved in a change in progress, or it is stigmatized, used by a man or a woman, etc. In Chapter Four, I review several counterexamples to Labov’s principle. As we might expect, all of the cases discussed in Chapter Four in which men lead a phonetic-level change that are below the level of social awareness (and therefore counterexamples to Labov’s claim) involve men producing more of a phonetic-level engendered variant that is associated with masculinity. As such, all of these counterexamples are adequately dealt with by the engendered–nongendered distinction. The end result is that the predictive power of Labov’s principle greatly increases once we acknowledge that the principle is intended to describe nongendered variation.

These points about engendered variation do not just apply to anatomically-determined variation. They are also intended for behaviourally-determined variation, specifically the point that women produce sharper phonetic distinctions than men, as expressed in Gender Hypothesis 1. The combination of Gender Hypothesis 1 and LV&C Hypothesis 1 yields the following hypothesis on the role sex in sound change below the level of social awareness.

Language Variation and Change (LV&C) Hypothesis 2 (page 118)

Women lead men in linguistic changes that involve the maintenance of phonetic distinctiveness, while men lead women in linguistic change that involves the loss of phonetic distinctiveness, as long as there are not overriding factors (such as prestige) involved.

LV&C Hypothesis 2 builds solidly on Gender Hypothesis 1, and these two hypotheses further illuminate the role of sex in language variation and change. Their central role in this dissertation and contribution to our understanding of language variation and change is reflected in the title of dissertation. This contribution was demonstrated in Chapter Six. In that chapter, I empirically demonstrated the link between vowel distinctiveness and vowel mergers by showing that speakers who lead vowel mergers in North American English also have a less dispersed vowel space. This leads to an explanation of why men lead mergers: men produce vowels that are less spread out than vowels produced by
women in general, and therefore men are already inclined towards vowels merging when vowel changes occur.

The significance of the engendered versus nongendered distinction is particularly relevant when it comes to sociolinguistic theories of language change. Language change, particularly change below the level of social awareness, is something we still understand very little about. As a case in point, consider the following discussion of why women lead language change published in *Psychology Today Magazine* (Aaronson 2005):

Why are women ahead of the curve? They tend to communicate more cooperatively than men, some researchers say, and thus may pick up others’ habits more quickly. Women, on average, also have stronger verbal skills than men. Some experts say women are more attuned to language and its quirks, given their primary role in caring for children and teaching them to speak. Men lag behind, perhaps because they are reluctant to copy women.

Look at the number of explanations given: women communicate more cooperatively than men, have stronger verbal skills, are more attuned to language because they are responsible for teaching children to speak, and men are reluctant to copy women. While some of these explanations seem more probable than others, they are all plausible. However, all of them suffer from the same fallacy; they do not explain why women lead *most but not all* sound changes below the level of social awareness. If women really do lead sound change because they are more attuned to language, then why do women lag behind on some changes, such as isolated vowel mergers?

By contrast, phonetic distinctiveness as an explanation of why women lead language change not only makes predictions that are easily testable (as demonstrated in Chapter Six), but also accounts for why men lead some of the sound changes. Specifically, men lead sound changes that result in either an increase in an anatomically-determined *masculine* variant (such as an increase in voicing), or a behaviourally-determined *masculine* variant (such as a loss of phonetic distinctiveness). Women on the other hand lead the sound changes that result in either an increase in an anatomically-determined *feminine* variant (such as an increase in devoicing), or a behaviourally-determined *feminine* variant (such as the maintenance of phonetic distinctiveness). Women lead a majority of the phonetic-level sound changes that are taking place below the level of social awareness because most sound changes involve the maintenance of phonetic distinctiveness, a feminine variant.
7.2 From specifics to generalizations, and back to specifics

7.2.1 From specifics to generalizations: The first step

Sociolinguists, in our impassioned pursuit of social explanations of language variation and change, sometimes fail to spot the underlying theoretical relevance of specific sociolinguistic findings. Understanding the specific nuances of language use at the level of the individual or community will always be the first step to understanding linguistic universals. But we cannot stop there. At some point we must move beyond the specifics — beyond the observations, for instance, that lower-class speakers use a more harsh vocal setting in Edinburgh and that Swedish men produce longer unstressed vowels than Swedish women. "We have to know these things before we can know more," Chambers (2005:216) says in a similar context, and further adds, "Eventually, however, we have to know more." In this thesis, I have attempted to move beyond the specifics to principles and theories that explain the roles of sex and gender in language variation and change.

Moving beyond specifics is not as straightforward in sociolinguistics as it is in the other subdisciplines of linguistics. In the sociolinguistics enterprise we have cultural and social factors to take into account as well as linguistic ones. Cultural considerations may cloud our vision, blinding us from seeing the general patterns. For example, consider the comment made by Labov, who is truly visionary when it comes to making observations about language, on the role of gender in language change (Labov 2001a:263): “Everyone agrees that gender is a social factor — language is not differentiated by the biological aspects of sex differences.” He is surely right in saying that everyone agrees that gender is a social factor. But he is surely wrong when he says that language is not differentiated by the biological aspects of sex differences. After all, there are some aspects of language — albeit, very few, but they exist nonetheless — that are differentiated by the biological aspects of sex differences. They include fundamental frequency, harmonics, and sibilant peak frequency, to name a few. Clearly Labov’s statement is too strong.

Similarly, Eckert and McConnell-Ginet’s (1992:462) appeal to scholars to abandon the assumption that linguistic manifestations of social gender are the same across communities is too strong. It is indisputably correct for most — but not all — linguistic manifestations of social gender. The assumption that linguistic manifestations of social gender are the same across communities must hold for anatomically-determined
phonetic-level engendered variation. How could a high frequency sibilant ever be
associated with machismo and toughness given the anatomical bias of female vocal tracts
towards the production of higher frequency sibilants than males? The same holds for
behaviourally-determined phonetic-level engendered variation. If we cannot investigate
consistency across communities, we are doomed to community-specific research.
Without that first step, we will never come to understand phonetic-level language
variation and change.

7.2.2  From generalization back to specifics: The next step

The research presented here has led us one small step closer to understanding phonetic-
level language change. The next step is to go from generalizations back to specifics, and
re-examine the phonetic distinctiveness of individuals in an attempt to understand better
who exactly instigates language change. For example, let us reconsider Chambers’s
discussion (2003:110–4) of “insiders,” the leaders of sound change, in light of phonetic
distinctiveness.

Chambers observes that “[t]here appears to be a tendency for individuals who are at
the centre of their social group to run ahead of the group in their use of salient markers”
(p.110). He illustrates this tendency with case studies from Boston (Fischer 1958),
Toronto (Chambers 1984), and Tokyo (Sibata 1960). The Tokyo study tracks the
dissemination of made-up words among a class of primary school children. Sibata
created a list of the social characteristics of the children that played the most active role
in disseminating words. Those children were described by Sibata as:

- cheerful
- popular
- active in social life in the class
- scholastically superior
- more emotional than reasonable
- likely to have elder siblings

Chambers points out that this list of characteristics describes the leaders of the other two
case studies very well. Of these characteristics, I want to consider more closely the trait “more emotional than reasonable,” as it connects us to phonetic distinctiveness.

Phonetic distinctiveness is an internal (i.e., non-social) motivation for sound change that is linked to clarity of speech in a technical, phonetically-defined sense. In the acknowledgements at the beginning of the dissertation, I briefly reviewed another internal motivation for sound change that was also linked to clarity of speech — prosodic prominence. Jacewicz et al. (2006) found that vowels produced with a high degree of prosodic prominence are shifted from non-prominent vowels in ways that correspond to Labov’s principles of chain shifts. Their study was motivated by Labov’s observation (2001a:439–42) that Carol Myers (a pseudonym for a speaker recorded in a variety of settings) produced more advanced vowel shifts in her emotional speech in comparison to her more business-like speech. Here is how the researchers view the link between emotional speech and prosody (p.289):

In short, we see emotional emphasis as one of the correlates of hyperarticulation (Lindblom 1990), which causes changes to acoustic vowel characteristics similar to what we find in more prosodically prominent positions. As a result of these acoustic changes, vowels spoken more emphatically stand out from the continuously varying speech stream and provide listeners with more perceptual cues than vowels uttered less clearly.

Vowels produced with a high level of prosodic prominence are more shifted in the direction of ongoing vowel changes than vowels produced with less prosodic prominence. As well, emotional speech results in more prosodically-prominent vowels. Lastly, vowels spoken more emphatically are clearer than less emphatic vowels. The implication here is that people who often speak with a lot of emotion produce more instances of vowels shifted in the direction of ongoing vowel changes, that is, they lead the change. They also produce vowels that are easier to perceive. It turns out that “emotional” is one of the characteristics of insiders listed by Sibata.

Chambers’s identification of the characteristics of insiders is based on language variation that is salient, i.e., above the level of social awareness. But it may very well be that these insiders are also leading changes below the level of social awareness, given their tendency towards emotional speech. This can be tested by taking several measures of phonetic distinctiveness of the speech of the leaders and comparing them to the speech of the non-leaders. These empirical tests lie in the future, but no doubt we will find that
these cheerful and popular speakers, these insiders of language change, also speak the most clearly, and produce phonological contrasts with greater phonetic distinctiveness than the non-leaders.

I began this dissertation by making a distinction between nongendered and engendered variation, as illustrated in Figure 1.1. The four boxes in the figure represent the following types of variation: nongendered segmental variation, nongendered suprasegmental variation, engendered segmental variation, and engendered suprasegmental variation. Sociolinguists and sociophoneticians have worked most comfortably within the box that represents segmental nongendered variation. I hope that my dissertation demonstrates that there can also be empirical rewards for thinking outside that box.
Appendix A: List of Hypotheses

This appendix contains a list of the phonetic-level engendered variation hypotheses, and the page reference for where in the dissertation they were introduced.

*Characteristics of phonetic-level engendered variation (pages 59, 79)*

Phonetic-level engendered variation displays the following characteristics:

- when enough speakers are looked at, sex-based linguistic patterns emerge
- these patterns of engendered variation tend to be consistent across different speech communities
- many of these patterns are grounded in anatomical sex differences in the vocal tract
- other patterns are clearly not grounded in anatomical sex differences in the vocal tract, and therefore are behavioural
- the extent of the sex differentiation seen in the engendered variation varies from one speech group / community to the next, but the direction the sexes pattern tends to remain the same
- however, there are individuals or groups who for social reasons pattern more like the opposite sex

*Gender Hypothesis 1 (page 84)*

The phonetic distinctiveness of the speech of women is greater than that of the speech of men.

Phonetic distinctiveness is defined as the acoustic and / or temporal distinction between the phonetic realizations of two related phonological constituents. The relationship between the phonological constituents can be either paradigmatic, syntagmatic, or both.
Gender Hypothesis 2 (page 86)

Phonetic distinctiveness indexes social gender.

Social Class Hypothesis 1 (page 92)

Working-class speakers use more masculine variants than middle-class speakers, while middle-class speakers use more feminine variants than working-class speakers.

Social Class Hypothesis 2 (page 93)

Working-class speakers show greater sex differentiation for both masculine and feminine engendered variants than middle-class speakers.

Gender Style Hypothesis 1 (page 106)

There is engendered variation both above and below the level of gender awareness. Phonetic-level engendered variation below the level of gender awareness does not participate in the dynamic construction of gender.

Gender Style Hypothesis 2 (page 107)

Women use masculine variation that is above the level of gender awareness more often than men use feminine variation that is above the level of gender awareness.

LV&C Hypothesis 1a (page 109)

Men use a masculine variant more frequently than women regardless of whether the variation is stable sociolinguistic stratification or a change in progress, as long as it is void of prestige.
**LV&C Hypothesis 1b (page 109)**

Women use a feminine variant more frequently than men regardless of whether the variation is stable sociolinguistic stratification or a change in progress, as long as it is void of prestige.

**LV&C Hypothesis 2 (page 118)**

Women lead men in linguistic changes that involve the maintenance of phonetic distinctiveness, while men lead women in linguistic change that involves the loss of phonetic distinctiveness, as long as there are not overriding factors (such as prestige) involved.
Appendix B: Transcriptions of Speech Samples

DJ One: Country Dan

Dan: Alabama here on XM Ten America. That one goin’ out to the Canadian Midnight Hauler who happens to be in Indiana, makin’ his way back to Canada, Toronto to be exact. Have a safe trip there Canadian Midnight Hauler. Have a safe trip, eh? Hahaha. Alright, now we’ve got one we are going to send out to Joe Lafantene in Louisberg, Florida, listenin’ to us on the computer. He says he likes those good old songs from Big Al Downing.

Dan: XM ten America. Buck Owens and the Bukaroos there with Rollin’ in my Sweet Baby’s Arms. Gonna send that out to John in Rockaway, New Jersey. Road Hazard emailed me. He tells me that his son, Road Puppy just got his C D L, and they have teamed up together, and asked for a song that J W played just before I got on the air, so I am going to do a little substitute here for you Road Hazard, here is Red Solvine and his classic called Giddy-up Go, Daddy! Giddy-up Go!

DJ Two: Ted Kelly

Ted: It’s our third year with the Brits, we are going to be there again this coming February, and also UPOP is gonna be bringing you the UPOP sessions at Abbey Road, between the Brits and the N M E awards. Some exciting things happening for the beginning of 2007 for the world’s pop music channel. It’s Ted Kelly, the World Party. Coast to coast in America. We’re the one place on earth that brings together the world’s biggest pop stars. From Turkey, he is one of the biggest stars on earth. It’s Tarkan.

Ted: Music mogul Russel Simons from Def Jam Music, and Def Jam Records, and Def Comedy, and Def Poetry. He is using hip hop as a medium to register what he hopes to be ten thousand new voters, in my home town of Philadelphia, Pennsylvania, he announced today. He said the non-partisan voter registration effort will be spearheaded through the Hiphop Summit Action Network, the H S A N, co-founded by Simons and Dr. Benjamin Travis, and the effort is being supported by the U S representative Shakapata, Philadelphia major John F. Strud, a good guy, and Governor Ed Rendal.
DJ Three: Lars Fredericksen

Lars: Then a great band Class Criminoly, and that was a little track off the old Scars and Upstarts, came out in 2002, Dwayne Peters compilation. There was a song called I’m A Junkie, Cras- Class Criminoly, good friends of ours obviously-

Gordon: Fuck yeah!

Lars: We met them going to the communist social, I should say, with Marco, come see us play. At the communist social.

Lars: What you heard, kicked it all off with Mad Sin, their Survival of the Sickest record. That was a track called Conquer the World. And then probably one of the greatest bands I’ve ever drink beer to, easy, ever drink beer to, fuckin’ did speed, acid, smoked fuckin’ weed.

Lars: Lars Fredericksen on a hot summer day, let me tell you what you heard. Kicked it off with Rancid, Indestructible Jango, Rancid Radio, might as well play Rancid, right? You know what I mean, and then the mighty Joe Strummer and the Mescolaros, off their last record, Street Core. A little Bob Marley song. It’s probably one of those songs that would be kinda hard to cover. But then again, it is Joe Strummer.

DJ Four: Chris Jericho

Chris: Good to see him back on the road though. He’s hooked himself up with John 5, who you may remember from Marilyn Manson, and also from David Lee Roth, who we spoke about earlier, gettin’ kicked off the air waves. Roth sacked. I love that word, a good sacking was had. Roth has been unceremoniously removed. But he did play with John 5 way back in the D L R band way back, if you heard the song Slam Duck which we played a few times, that was John 5 on guitars, and John 5 also played on the new Paul Stanley solo record, so John is kinda becoming the poster boy for rock and roll bands, kinda the gun for hire. And Educated Horses record, it reminds me a lot of Rob’s last record. Rob to me is kind of like a Rage Against the Machine type of artist. He puts out a record with two or three really kick ass songs, and then the rest just kind of blends in. But it gives him an excuse to get back out on the road and play some of the White Zombie songs, which is what everyone really wants.
DJ Five: Martin Goldsmith

Martin: One of the very last string quartets by the man that more or less invented the form. Those were the members of the Jerusalem Quartet. I’m Martin Goldsmith with the music for you here on XM Classics, we are channel 110, at XM Satellite Radio. As you might imagine this year of 2006, the 250th anniversary of the birth of Wolfgang Amadeus Mozart, has seen the release of lots of Mozart compact discs, both new and reissues. Here is a new recording of the violin concerto Number Five in A major, known as Mozart’s Turkish concerto. This comes from a recent two CD Set released by Deutsche Gramophone, in time for Mozart’s birthday back in January. Zophie Modor plays the violin and also conducts the London Philharmonic Orchestra. This is Mozart’s violin concerto Number Five, the Turkish concerto, here on XM Classics.

DJ Six: Coolguy

Coolguy: Forty-Two, Liquid Metal. We’re gonna do Fear Factory, Remembering Never, also Colenis Circle and Nuclear Assault, right now though, this is Scar Symmetry, great stuff off Pitch Black Progress CD. It’s Calculate the Apocalypse.

Coolguy: Forty-Two, Liquid metal, that was Lamb of God, we’re gonna do Hymns, and also Rage, some people, some people don’t think Rage Against The Machine belongs on Liquid Metal, I ain’t sayin’ nothin’, we’re gonna do Killing In the Name of, off their Hymns, we’re also gonna do Misery and Slayer. Right now though, it’s Black Sabbath off the Sabbath, Bloody Sabbath CD, right here.

Coolguy: It’s Forty-Two, Liquid Metal, Life Once Lost, also Hate Breed, Mortal Trees, let’s take a call. What’s happenin’?

Caller: What’s up, Coolguy?

Coolguy: Hey, tiger, what are you doin’?

Caller: Not a lot, man.

Coolguy: You ever think about fuckin’ a snowman? To past the time?

Caller: Hehehe.

Coolguy: This ain’t no joke man, I’m tellin’ you it-

Caller: What you all got goin’ on? Gay Night tonight or somethin’, dude?

Coolguy: Fuckin’ a snowman’s not gay, what are talkin’ about?
Caller: [grunts in disagreement]
Coolguy: Huh?
Caller: I was wonderin’ if you could play a song for me, bro?
Coolguy: What do you need? Frosty?
Caller: No, sir. Not tonight.
Coolguy: Alright, maybe tomorrow. What do you need?
Caller: How about Nothing Face?
Coolguy: I don’t think I’ve played any Nothing Face in a while. Alright.

DJ Seven: Paul Bachman

Paul: XM Classics. Well, if you are familiar at all with Joachim Raff, it is most likely because of his symphonic output. He is the much underrated late romantic symphonic composer. I am a big fan of his seasonal symphonies and his Lenoir Symphony is a real favourite of mine. And much more, Martin Goldsmith, my partner here at XM classics, even more of a Raff fan, and you know what, you can find out more about X online. There is a Raff fan site. Yeah that’s right. It is r a f f dot org. So you could check out for more information for yourself. Or you could always visit us online. Plenty of information to be found on the XM satellite radio website, XM radio dot com. Next, music by Mozart, another pretty new recording, as was the last one by Joachim Raff, that was new for 2006. As is this, from the West Eastern Dilon Orchestra, conducted by Daniel Barenboim. Mozart, the Sinfonia Concertante in E flat, for oboe, clarinet, bassoon, and orchestra, on XM One Ten.

DJ Eight: J.W.

J.W.: We had to play Wolf Creek Pass because Bill hadn’t heard it on America before, and been with us like forever. From the beginning. There you go Bill. Thanks for the calls, Smooth Move, Dick Dog, Checkers, we got one down. And Rooster. Johnny Kermit, and Wild Streak, and Jim wants some Conway on XM ten.

J.W.: That’s The Way Love Goes, on America. It’s XM Ten, I’m J.W. in the Wood Shed, the rented one. I was lookin’ over a list here, of on this day in History, the top records and stuff like that, and I picked out three of them. I don’t know. It must be my lucky day today. We are
going to play three in a row, from, let’s see, 1956, 1964, and 1980. We’re going to start with 1980. It was number one on this day back then. Could be the greatest country song of all time, George Jones on XM ten.
Appendix C: Zero-crossing Rate PRAAT Script

A description of the script

The following page is the section of the PRAAT script used to measure the zero-crossing rate of a fricative. The script steps the cursor through the speech signal looking for zero crossings using PRAAT’s Get Nearest Zero Crossing function. This function returns the time location of the zero crossing closest to the cursor, regardless of whether the crossing is in front of or behind the cursor. When a new zero crossing is found, the crossing counter is incremented, the location of the crossing along the time axis is recorded, and the cursor is advanced one step. We then look for the closest crossing again. It is possible that the closest crossing to the cursor after it has been stepped forward is still the crossing that was previously found, so the time locations of the new crossing and the previously found crossing are compared. If the locations are the same, then just the cursor is advanced, and the process is repeated. If the locations are different, then the counter is incremented and the time location of the new crossing is recorded before the cursor is advanced.

Note that the step size is set at 0.000005 seconds, or 200,000 steps per second. This is ten times the resolution of the highest frequency perceived by the human ear (about 20,000 Hz). The highest zero crossing rates recorded were just over 10,000 crossings per second. Such a high resolution guarantees that we are not skipping over crossings, which would result in an under-representation of the true crossing rate.
The script:

```
step = 0.000005
total_x = 0
zxt = Get nearest zero crossing... 'current_t'; a PRAAT function
old_zxt = zxt; remember where the zero crossing is

while zxt <> undefined ; returns “undefined” if no crossing found
    current_t = current_t + step
    zxt = Get nearest zero crossing... 'current_t'
    if zxt = old_zxt
        current_t = current_t + step
    else
        current_t = current_t + step
        total_x = total_x + 1
        old_zxt = zxt
    endif
    if total_x > 20000 ; check to see if crossing rate too high
        exit SAFETY ABORT: POSSIBLE LABEL ERROR?
    endif
endwhile

select Sound untitled
Remove

printline total zero crossings in = 'total_x'
zero_x_rate = total_x / duration
printline zero crossing rate = 'zero_x_rate:2' crossings / s
```
Appendix D: Merger Cross Tabulations

The following pages contain the cross tabulations for the twelve mergers used to test for a link between vowel distinctiveness and merger category. Each merger was cross tabulated for merger category (“merged,” “close” or “distinct”) and region. Each of the regions was then assigned an index score, representing the degree of merger in that region. The index score was calculated by assigning the responses in the merged category a weight of two, the responses in the close category a weight of one and the responses in the distinct category a weight of zero. These weights were then summed and divided by the total number of responses. This resulted in a number between zero and two, which was then multiplied by 100 for ease of presentation. Following is the calculation for the index as a formula:

\[
\text{index} = \frac{( \# \text{ of MERGE} \times 2 + \# \text{ of CLOSE} )}{N}
\]

A value of 200 represents a complete merger in that region, whereas a value of zero represents a complete distinction.

The index scores were then used to decide which regions to include in the statistics calculations in Chapter Six. Regions that showed no or limited variation among the speakers were excluded from the calculations. A region was considered to show no or limited variation if its index score was between 0 and 24, or between 176 and 200. Regions with index scores between 25 and 175 were included in the calculations.
<table>
<thead>
<tr>
<th>Merger: DOLL- TALL</th>
<th></th>
<th>Merger: DON- DAWN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td><strong>Canada</strong></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>8</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Other Canada</td>
<td>24</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td><strong>Eastern New England</strong></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td><strong>Midland</strong></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>43</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td><strong>Mid-Atlantic</strong></td>
<td></td>
<td><strong>Mid-Atlantic</strong></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>30</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>St. Louis</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other North</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td><strong>South</strong></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Other South</td>
<td>54</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td><strong>Southeast</strong></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Transitional</strong></td>
<td></td>
<td><strong>Transitional</strong></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>48</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>Western Penn.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>8</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>
### Merger: NOT–NAUGHT

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>173</td>
</tr>
<tr>
<td>Other Canada</td>
<td>24</td>
<td>19</td>
<td>3</td>
<td>2</td>
<td>171</td>
</tr>
<tr>
<td>Eastern New England</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>150</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Midland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>133</td>
</tr>
<tr>
<td>Other Midland</td>
<td>61</td>
<td>29</td>
<td>10</td>
<td>22</td>
<td>111</td>
</tr>
<tr>
<td>Mid-Atlantic North</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>62</td>
<td>1</td>
<td>7</td>
<td>54</td>
<td>15</td>
</tr>
<tr>
<td>St. Louis</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>75</td>
</tr>
<tr>
<td>Other North</td>
<td>38</td>
<td>10</td>
<td>7</td>
<td>21</td>
<td>71</td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>150</td>
</tr>
<tr>
<td>Other South</td>
<td>64</td>
<td>16</td>
<td>4</td>
<td>44</td>
<td>56</td>
</tr>
<tr>
<td>Southeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>133</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Transitional West</td>
<td>26</td>
<td>15</td>
<td>2</td>
<td>9</td>
<td>123</td>
</tr>
<tr>
<td>West</td>
<td>48</td>
<td>42</td>
<td>4</td>
<td>2</td>
<td>183</td>
</tr>
</tbody>
</table>

### Merger: SOD–SAWED

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other Canada</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Eastern New England</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Providence</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Midland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other Midland</td>
<td>31</td>
<td>13</td>
<td>5</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td>Mid-Atlantic North</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>St. Louis</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>West. New England</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>133</td>
</tr>
<tr>
<td>Other North</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>67</td>
</tr>
<tr>
<td>New York City</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Texas South</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other South</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>Southeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Florida</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transitional West</td>
<td>9</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>144</td>
</tr>
<tr>
<td>West</td>
<td>36</td>
<td>30</td>
<td>3</td>
<td>3</td>
<td>175</td>
</tr>
<tr>
<td>Western Penn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>140</td>
</tr>
</tbody>
</table>
### Merger: HILL–HEEL

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Canada</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>59</td>
<td>7</td>
<td>4</td>
<td>48</td>
<td>31</td>
</tr>
<tr>
<td><strong>Mid-Atlantic North</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>60</td>
<td>0</td>
<td>2</td>
<td>58</td>
<td>3</td>
</tr>
<tr>
<td>St. Louis</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Other North</td>
<td>37</td>
<td>2</td>
<td>3</td>
<td>32</td>
<td>19</td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>11</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>145</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>87</td>
</tr>
<tr>
<td>Other South</td>
<td>64</td>
<td>16</td>
<td>9</td>
<td>39</td>
<td>64</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Transitional West</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>43</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>20</td>
</tr>
</tbody>
</table>

### Merger: HOCK–HAWK

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Canada</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>60</td>
<td>21</td>
<td>11</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td><strong>Mid-Atlantic North</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>60</td>
<td>2</td>
<td>2</td>
<td>56</td>
<td>10</td>
</tr>
<tr>
<td>St. Louis</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>67</td>
</tr>
<tr>
<td>Other North</td>
<td>38</td>
<td>9</td>
<td>7</td>
<td>22</td>
<td>66</td>
</tr>
<tr>
<td>New York City</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Other South</td>
<td>54</td>
<td>7</td>
<td>3</td>
<td>44</td>
<td>31</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>133</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td><strong>Transitional West</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>10</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>140</td>
</tr>
</tbody>
</table>
## Merger: BELL–BAIL

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other Canada</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Providence</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>55</td>
<td>5</td>
<td>4</td>
<td>46</td>
<td>25</td>
</tr>
<tr>
<td><strong>Mid-Atlantic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>58</td>
<td>1</td>
<td>2</td>
<td>55</td>
<td>7</td>
</tr>
<tr>
<td>St. Louis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>West. New England</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other North</td>
<td>38</td>
<td>2</td>
<td>4</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>New York City</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>11</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>145</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>Other South</td>
<td>62</td>
<td>16</td>
<td>10</td>
<td>36</td>
<td>68</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Transitional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>47</td>
<td>6</td>
<td>5</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td><strong>Western Penn.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>20</td>
</tr>
</tbody>
</table>

## Merger: FULL–FOOL

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Other Canada</td>
<td>24</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>60</td>
<td>7</td>
<td>3</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td><strong>Mid-Atlantic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>60</td>
<td>1</td>
<td>5</td>
<td>54</td>
<td>12</td>
</tr>
<tr>
<td>St. Louis</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>Other North</td>
<td>38</td>
<td>1</td>
<td>3</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>11</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>63</td>
</tr>
<tr>
<td>Other South</td>
<td>64</td>
<td>7</td>
<td>11</td>
<td>46</td>
<td>39</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>67</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Transitional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>48</td>
<td>5</td>
<td>9</td>
<td>34</td>
<td>40</td>
</tr>
<tr>
<td><strong>Western Penn.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>157</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>120</td>
</tr>
</tbody>
</table>
## Merger: MERRY–MARY

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>150</td>
</tr>
<tr>
<td>Other Canada</td>
<td>24</td>
<td>18</td>
<td>2</td>
<td>4</td>
<td>158</td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Providence</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other Midland</td>
<td>18</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>194</td>
</tr>
<tr>
<td><strong>Mid-Atlantic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>31</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>197</td>
</tr>
<tr>
<td>St. Louis</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>West. New England</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>133</td>
</tr>
<tr>
<td>Other North</td>
<td>31</td>
<td>29</td>
<td>2</td>
<td>0</td>
<td>194</td>
</tr>
<tr>
<td><strong>New York City</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>163</td>
</tr>
<tr>
<td><strong>Other South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other South</td>
<td>58</td>
<td>33</td>
<td>2</td>
<td>23</td>
<td>117</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td><strong>Transitional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitional</td>
<td>16</td>
<td>11</td>
<td>0</td>
<td>5</td>
<td>138</td>
</tr>
<tr>
<td><strong>West</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Penn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>160</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
</tbody>
</table>

## Merger: PIN–PEN

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>Other Canada</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td><strong>Eastern New England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td><strong>Midland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>60</td>
<td>28</td>
<td>8</td>
<td>24</td>
<td>107</td>
</tr>
<tr>
<td><strong>Mid-Atlantic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>56</td>
<td>3</td>
<td>7</td>
<td>46</td>
<td>23</td>
</tr>
<tr>
<td>St. Louis</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>63</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Other North</td>
<td>36</td>
<td>3</td>
<td>6</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td><strong>New York City</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td><strong>South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>200</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>138</td>
</tr>
<tr>
<td><strong>Other South</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other South</td>
<td>64</td>
<td>50</td>
<td>5</td>
<td>9</td>
<td>164</td>
</tr>
<tr>
<td><strong>Southeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>167</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>160</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td><strong>Transitional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitional</td>
<td>26</td>
<td>4</td>
<td>3</td>
<td>19</td>
<td>42</td>
</tr>
<tr>
<td><strong>West</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Penn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>57</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>
## Merger: HOARSE–HORSE

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Canada</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern New England</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Providence</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other ENE</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Midland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other Midland</td>
<td>59</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>0</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>North</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>59</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>0</td>
</tr>
<tr>
<td>St. Louis</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>71</td>
</tr>
<tr>
<td>West. New England</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Other North</td>
<td>36</td>
<td>0</td>
<td>1</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>New York City</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Other South</td>
<td>61</td>
<td>1</td>
<td>1</td>
<td>59</td>
<td>5</td>
</tr>
<tr>
<td>Southeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Transitional</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>West</td>
<td>47</td>
<td>1</td>
<td>1</td>
<td>45</td>
<td>6</td>
</tr>
<tr>
<td>Western Penn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

## Merger: MERRY–MARRY

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Merge</th>
<th>Close</th>
<th>Distinct</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic Provinces</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>175</td>
</tr>
<tr>
<td>Other Canada</td>
<td>24</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>183</td>
</tr>
<tr>
<td>Eastern New England</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Providence</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Other ENE</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Midland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cincinnati</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other Midland</td>
<td>18</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>194</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>108</td>
</tr>
<tr>
<td>North</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland North</td>
<td>31</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>197</td>
</tr>
<tr>
<td>St. Louis</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>West. New England</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>167</td>
</tr>
<tr>
<td>Other North</td>
<td>32</td>
<td>31</td>
<td>1</td>
<td>0</td>
<td>197</td>
</tr>
<tr>
<td>New York City</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inland South</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td>Texas South</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>179</td>
</tr>
<tr>
<td>Other South</td>
<td>58</td>
<td>50</td>
<td>4</td>
<td>4</td>
<td>179</td>
</tr>
<tr>
<td>Southeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>167</td>
</tr>
<tr>
<td>Florida</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other Southeast</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Transitional</td>
<td>16</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>144</td>
</tr>
<tr>
<td>West</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Western Penn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Other West. Penn.</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
</tbody>
</table>
References


Biemans, Monique. 2000. Gender variation in voice quality. PhD dissertation, University of Nijmegen,
Nijmegen, Netherlands.


Docherty, Gerard J., Foulkes, Paul, Milroy, James, Milroy, Lesley, and Walshaw, David. 1997. Descriptive


Archives of Otolaryngology 92, 379–382.


Lau, Yiu-Kei and Chan, Chok-Ki. 1985. Speech recognition based on zero crossing rate and energy. *IEEE
Transactions on Audio and Electroacoustics 33, 320–323.


REFERENCES


Ryalls, John, Zipprer, Allison, and Baldauff, Penelope. 1997. A preliminary investigation of the effects of


Thomas, Erik R. 2001. *An Acoustic Analysis of Vowel Variation in New World English*. Durham, NC:
REFERENCES

Duke University Press for the American Dialect Society.


Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. _Language in Society_ 1, 179–195.


Wolf, Clara and Jiménez, Elena. 1979. A sound change in progress: Devoicing of Buenos Aires */h/.

Unpublished manuscript.


