



**LIBRARY**  
**Michigan State**  
**University**

**This is to certify that the  
dissertation entitled**

## INTERDISCIPLINARY PERSPECTIVES ON THE NORTHERN CITIES CHAIN SHIFT

presented by

## Barthomiej Plichta

has been accepted towards fulfillment  
of the requirements for the

**Ph.D**                      **degree in**                      **Linguistics and Germanic,  
Slavic Asian and African  
Languages**

For Release 

Major Professor's Signature

12/2/04

**Date**

**MSU is an Affirmative Action/Equal Opportunity Institution**

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
06-28-2009		

**INTERDISCIPLINARY PERSPECTIVES ON THE NORHTERN CITIES CHAIN  
SHIFT**

**By  
Bartłomiej Plichta**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Linguistics and Germanic, Slavic, Asian and African Languages**

**2004**



## ABSTRACT

### INTERDISCIPLINARY PERSPECTIVES ON THE NORTHERN CITIES CHAIN SHIFT

By

Bartłomiej Plichta

This dissertation concerns sociophonetic aspects of the Northern Cities Chain Shift (NCCS). An interdisciplinary research paradigm is proposed whereby certain articulatory-phonetic and perceptual processes are integrated into a hearer-mediated model of sound change. Two studies are presented – one is speech production, and one in speech perception. The production study demonstrates that vowel nasalization is distributed along the NCCS isoglosses of sex and region. The perception study shows that sociophonetic information is utilized in speech perception. The two studies come together in an argument that the NCCS vowel space is being reshaped by micro-level articulatory-perceptual processes negotiated by the speaker, hearer, and the speech community.

## ACKNOWLEDGEMENTS

I am grateful to Professors Dennis Preston and Brad Rakerd who have inspired and supervised this work with their tremendous expertise and professionalism.

I would like to thank Professor Dennis Preston and Professor David Prestel for their support in the development of the Macintosh version of *Akustyk*.

This work has been made possible in part by the National Science Foundation Award #353162.

Finally, I would like to express my gratitude to open-source and scholarly communities whose work has made this project possible (in alphabetical order): Electronic Metastructure for Endangered Languages Data (EMELD.org), Linguistlist.org, OpenOffice.org, Praat, Scalable Vector Graphics (SVG), Содиподи, Tcl/Tk.

## PREFACE

*This dissertation has been formatted in compliance with the Michigan State University Graduate School dissertation formatting guidelines.*

*Statistics reports and citations have been formatted in compliance with the 5<sup>th</sup> Edition APA Publication Manual.*

*Figures on pages: xiii, 7, 14, 25, 27,28,33 41, 42, 47, 47, 48, 49, 50, 52, 54, 63, 84, 86, 87, 91, 97, 97, 100, 122 were generated by Akustyk (Plichta, 2004) and the EPS rendering library by Boersma & Weenink (2002).*

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
PREFACE.....	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES.....	viii
LIST OF FIGURES .....	ix
KEY TO SYMBOLS AND ABBREVIATIONS .....	xiii
CHAPTER 1 INTRODUCTION .....	1
1.1 THE GOALS OF THIS DISSERTATION .....	1
1.2 VARIATIONIST SOCIOPHONETICS.....	2
1.2.1 This dissertation as a work in variationist sociophonetics.....	3
1.2.2 Current state of the field of sociophonetics .....	3
1.3 SOCIOPHONETIC STUDIES OF SPEECH PRODUCTION .....	4
1.3.1 Quantitative studies of consonantal variation.....	4
1.3.2 Quantitative studies of vocalic variation.....	6
1.3.3 Sociophonetic accounts of suprasegmental features.....	8
1.4 SOCIOPHONETIC STUDIES OF SPEECH PERCEPTION.....	9
1.4.1 Perceptual dialectology.....	9
1.4.2 Experimental studies of speech perception.....	10
1.4.3 Needed research in perceptual sociophonetics .....	16
1.4.4 Speech Science –similar questions, different methodologies.....	17
1.5 ORGANIZATION.....	21
1.5.1 The speech production study .....	21
1.5.2 The speech perception study.....	21
1.5.3 Comments and conclusions.....	22
CHAPTER 2 VOWEL NASALIZATION AND NCCS .....	23
2.1 INTRODUCTION .....	23
2.2 FRONTING AND RAISING.....	24
2.3 PROBLEMS WITH TRADITIONAL ACOUSTIC ANALYSIS OF VOWELS .....	25
2.3.1 Data acquisition problems.....	25
2.3.2 Data acquisition with a standard analog Marantz recorder.....	26
2.3.3 Digital recording with a MiniDisc player .....	29
2.3.4 24-bit digital recording with a close-talking microphone.....	31

2

2

2

2

2

2

0

3

2.3.5	The role of the microphone.....	32
2.3.6	Different signal acquisition methods return different formant values .....	35
2.3.7	LPC analysis of nasalized vowels: Is /æ/ really raising? .....	41
2.4	VOWEL NASALIZATION .....	43
2.4.1	The velopharyngeal port .....	43
2.4.2	Oral and nasal formants .....	45
2.4.3	Spectral characteristics of synthetic nasalized vowels.....	46
2.4.4	Spectral characteristics of the oral ~ nasal contrast in Polish.....	47
2.4.5	LPC and nasalized vowels – evidence from Polish .....	49
2.5	VOWEL NASALIZATION IN MICHIGAN .....	51
2.5.1	Nasal formants appear in the spectrum of Lower Michigan vowels .....	51
2.5.2	Why a sociophonetic aerodynamic analysis of vowel nasalization? .....	52
2.5.3	Quantifying vowel nasalization for sociophonetic purposes .....	53
2.6	STUDY DESIGN AND METHODS .....	60
2.6.1	The goals.....	60
2.6.2	The subjects .....	61
2.6.3	Data collection .....	62
2.6.4	Why /i/ and /u/ were excluded from the study .....	63
2.6.5	Data processing.....	64
2.7	STATISTICAL ANALYSIS OF %N.....	65
2.7.1	%N - two-way analysis of variance .....	66
2.7.2	Follow-up test #1 – simple main effects tests .....	67
2.7.3	Follow-up test #2 – pairwise comparisons.....	67
2.7.4	Follow-up test #3 - interaction comparisons.....	68
2.7.5	Summary of the statistical analysis of % Nasalance.....	68
2.7.6	Is nasalization global or local?.....	69
2.8	STATISTICAL ANALYSIS OF A1-P1 .....	70
2.8.1	Summary of the spectral method .....	71
2.9	NASALIZATION AND NCCS.....	73
2.9.1	The <i>F2</i> of /a/ as an index of talker participation in NCCS .....	74
2.9.2	Investigating the correlation of %N and Bark-transformed <i>F2</i> .....	80
2.9.3	Summary .....	81
CHAPTER 3	PERCEPTIONS OF /a/-FRONTING .....	83
3.1	INTRODUCTION .....	83
3.1.1	Introduction to talker normalization .....	83

3.2	THE STUDY .....	89
3.2.1	Motivation for the study.....	89
3.2.2	Lower Michigan and the Upper Peninsula.....	89
3.2.3	Talker normalization across the LM and UP dialects.....	92
3.2.4	The subjects .....	92
3.2.5	The stimuli .....	93
3.2.6	The experiment .....	102
3.2.7	Analysis and results .....	106
3.2.8	Summary .....	114
CHAPTER 4	ON HEARER-MEDIATED SOUND CHANGE .....	116
4.1	HOW DO DIALECTS DIFFER? .....	116
4.1.1	Speaker-mediated sound change in dialectology.....	116
4.1.2	The need for a broader model of sound change.....	116
4.2	NASALIZATION, /a/-FRONTING, AND SOUND CHANGE.....	119
4.2.1	Articulatory variability and sound change.....	119
4.2.2	Vowel nasalization and perceived height .....	120
4.2.3	Hearer-mediated /a/-fronting .....	123
4.3	TOWARD A SOCIOPHONETIC MODEL OF HEARER-MEDIATED SOUND CHANGE .....	125
4.3.1	Phonological models of nasalization .....	125
4.3.2	A sociophonetic model .....	127
APPENDIX A	- WORDLIST USED IN CHAPTER 2 .....	130
APPENDIX B	- QUESTIONNAIRE (CHAPTER 2 and CHAPTER 3CHAPTER 2)....	131
APPENDIX C	- SUMMARY OF INDIVIDUAL CASES OF %N.....	132
APPENDIX D	- CONSENT FORM CHAPTER 2.....	133
APPENDIX E	- CONSENT FORM CHAPTER 3 .....	134
APPENDIX F	- WORDLIST USED IN CHAPTER 3 .....	135
BIBLIOGRAPHY	.....	136

T

T

T

T

Ta

Ta

Ta

Ta

Ta

Ta

Ta



## LIST OF TABLES

- Table 1. Summary of statistical results of the data acquisition test 38
- Table 2. Summary of simple effect tests for men and women across the three regions and for men and women for each region separately 67
- Table 3. Summary of simple effect tests for men and women across each region separately 68
- Table 4. Summary of three tetrad comparisons to evaluate whether the differences in %N means across the regions were the same or different for male and female respondents 68
- Table 5. Overall means and standard deviations for the parameters obtained with the Chen method 71
- Table 6. LM and UP respondents' normalized, mean  $F1$  and  $F2$  values in Hertz 92
- Table 7. Talker LM's and UP's mean  $F1$  and  $F2$  values in Hertz 100
- Table 8. Mean stimulus values by stimulus type and listener region 109
- Table 9. Wordlist used in studies described in CHAPTER 2 130
- Table 10. Summary of individual cases for %N 132
- Table 11. Wordlist used in the study in CHAPTER 3 135

## LIST OF FIGURES

- Figure 1. American English vowels talked about in this dissertation xiii
- Figure 2. Vowel quadrilateral with IPA-style phonetic symbols 7
- Figure 3. NCCS vowels and their movement within the two-dimensional  $F1/F2$  space 11
- Figure 4. Responses to the word "pop" (From Niedzielski (1999) with permission) 12
- Figure 5. Formant tracks of the 7 variants of the male pronunciation of the word "guide"  
14
- Figure 6. Mean responses to male and female voices 15
- Figure 7. /i/ and /ɪ/ and /p/ and /b/ identification and discrimination results (From Pisoni  
(1973) with permission) 19
- Figure 8. Simplified spectrograms of the stimuli used in the experiment From Fitch, et al  
(1980) 20
- Figure 9. The fronting of /a/ and the raising of /æ/ 25
- Figure 10. LPC of the vowel /i/ superimposed on the noise spectrum of the Marantz  
recorder 27
- Figure 11. Spectrum of 60 Hz hum and the vowel /a/ in "job" 28
- Figure 12 Waterfall plot of the word "hat" by an NCCS-influenced female talker 31
- Figure 13 Waterfall plot of the word "hat" by an NCCS-influenced female talker 32
- Figure 14. Low-end bias due to proximity effect of the AKG C-420 microphone 33
- Figure 15. Frequency response of Shure Beta 87a and Earthworks M30 35
- Figure 16. Box plot of overall between-subject comparisons of the data acquisition test 39
- Figure 17. LPC spectra of the /a/ vowel in "job" by a female talker with an NCCS-  
influenced vowel system acquired with by different methods 41
- Figure 18. Sample vowel systems of Detroit, MI females (from Labov et al. (1997) with  
permission) 42
- Figure 19. Schematic view of lowering (opening) of the velum during the production of  
nasalized vowels 44

Figure 20. Schematic view of the oral and nasal passages (of a male talker) involved during the production of nasalized vowels (based on Chen (1997)) 45

Figure 21 A theoretical model of the relationship between the area of the opening of the velopharyngeal port and the frequency and amplitude of oral and nasal formant for the vowel /æ/ (from Stevens (1998)) 46

Figure 22. Spectrum of the /æ/ vowel synthesized with the velopharyngeal port opening of 0 mm<sup>2</sup>, (left), and 7 mm<sup>2</sup>, (right) 47

Figure 23 Spectrum of the /æ/ vowel synthesized with the velopharyngeal port opening of 14 mm<sup>2</sup> (left) and 21 mm<sup>2</sup> (right) 47

Figure 24. Spectrogram of a Polish minimal pair /krɛta/~/krɛ̃ta/ 48

Figure 25. Spectral characteristics of the oral ~ nasal contrast in Polish 49

Figure 26. Comparison of LPC and FFT spectra of /ɛ/ and /ɛ̃/ 50

Figure 27 Examples of non-nasalized vowel spectra (left) and nasalized spectra (right.) of the vowel /æ/ in "back" 52

Figure 28. Waterfall plots of the word "back" of the same sample as in Figure 24 52

Figure 29. Oral and nasal prominences involved in Chen's method 54

Figure 30. Flowchart of *Akustyk*'s algorithm designed to automate the peak seeking process 56

Figure 31. The Rothenberg mask 59

Figure 32. Spectra from the oral and nasal channel of /a/ in "lot" 63

Figure 33. % Nasalance means by respondent region and sex 66

Figure 34. Male and female %N distribution patterns by vowel in non-nasal environments and for non-high vowels 70

Figure 35. Summary of *A1-P1* results obtained in the spectral method (higher *A1-P1* = lower nasalization) 72

Figure 36. Continuous % Nasalance levels for the words "dad" and "man" 73

Figure 37. Overview of *Akustyk*'s analysis tools 76

Figure 38. Cochleagram of /æ/ in "back" by a female Detroitter 79

- Figure 39. Scatter plot of Bark-transformed  $F2$  and Mean % N fitted around the regression line 81
- Figure 40. Differences in  $F1$  and  $F2$  due to VTL variability between men and women (from the Peterson and Barney corpus (1952)) 84
- Figure 41. Bark-transformed, discriminant plane of formant values from the Peterson and Barney corpus 86
- Figure 42. Bark-transformed, discriminant plane of formant values from a corpus of 26 adult NCCS speakers 87
- Figure 43. Lower Michigan and the Upper Peninsula 90
- Figure 44. Normalized, mean formant values of LM and UP participants in the study 91
- Figure 45.  $F1$  and  $F2$  tracks of the 7-step continuum of “sock-sack” used in the experiment 96
- Figure 46. Spectrographic images of the first step and last step of the /a/~ /æ/ continuum 97
- Figure 47. LPC spectra of the first and last step of the continuum 97
- Figure 48. Vowel systems of talkers LM and UP 100
- Figure 49 Finalized stimulus: a UP or LM precursor phrase with a synthesized target hVt or sVk word at the end 101
- Figure 50. Spectrogram of a carrier phrase fragment by Talker UP 102
- Figure 51. Psychometric functions of LM respondents to the stimuli with LM precursors and UP precursors 108
- Figure 52. Psychometric functions of UP respondents to the stimuli with LM precursors and UP precursors 108
- Figure 53. LM and UP responses as a function of the precursor phrase 111
- Figure 54. Summary of the word effect 112
- Figure 55. Precursor phrases 1 through 4 (P1-P4) - not significant in vowel identity judgments (From Rakerd and Plichta (2003) with permission) 113
- Figure 56. Normalized Principal Components plot of an NCCS population from Lower Michigan 122
- Figure 57. Summary of the Ohalan listener-mediated model of sound change 126

**Figure 58. Sociophonetic model of hearer-mediated sound change 128**

## KEY TO SYMBOLS AND ABBREVIATIONS

1. AAVE – African American Vernacular English (also known as AAE – African American English)
2. ATRAC – Adaptive TRansform Acoustic Coding
3. FFT – Fast Fourier Transform
4. LPC – Linear Predictive Coding.
5. NCCS – the Northern Cities Chain Shift (also known as NCS)
6. RMS – Root Mean Square
7. VOT – Voice Onset Time
8. VTL – vocal tract length
9. Vowel symbols with examples (Figure 1):

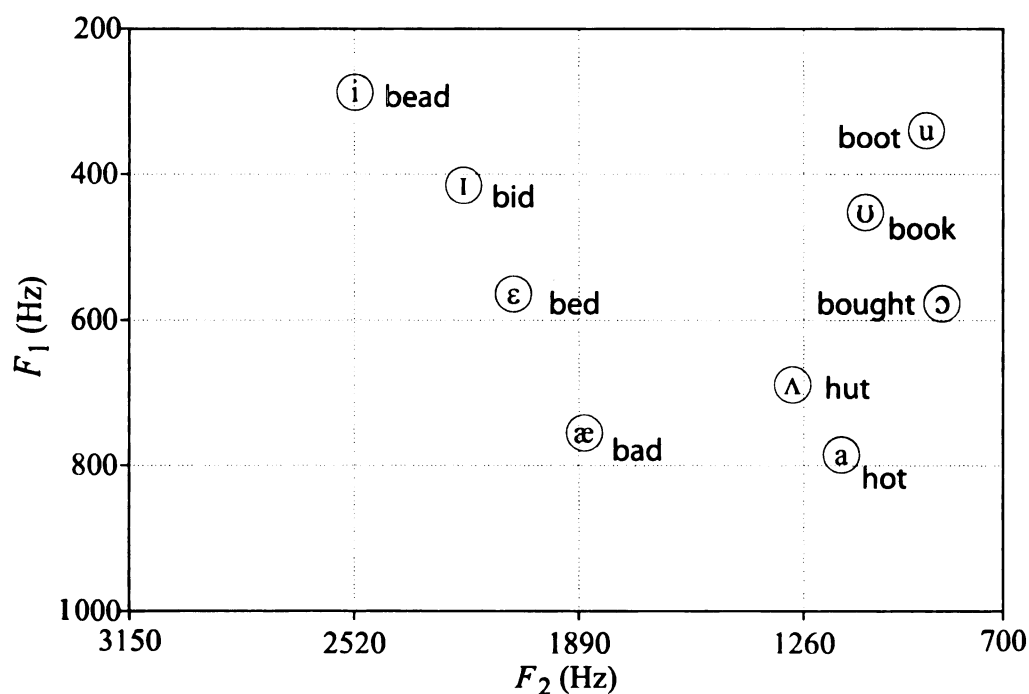


Figure 1. American English vowels talked about in this dissertation

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 THE GOALS OF THIS DISSERTATION**

Sociolinguists and dialectologists are generally interested in studying speech production as they trace the distribution of such phenomena as the Northern Cities Chain Shift (NCCS) through social and geographical space. More recently, they have also become concerned with speech perception as it enables them to investigate the effects of such sound changes on both comprehension and attitude and, perhaps even more centrally, as it allows them to look for the origins of sound change in speech perception (e.g., Thomas (2002)).

From both these points of view (production and perception), students of variation have increasingly relied on methodologies and instrumentation of related areas of scientific inquiry. Speech science, for example, offers a number of sophisticated practices involving computer-assisted acoustic analysis (in studies of speech production) and speech synthesis (in studies of speech perception). This dissertation includes two studies based on such practices. The first, an acoustic study of vowel nasalization in an on-going sound change (NCCS), suggests that previous accounts of NCCS can be broadened by an aerodynamic study of oral and nasal airflow, and that vowel nasalization plays a considerable role in the re-shaping of the NCCS-influenced vowel space. The second study provides a perceptual account the /a/~/æ/ category boundary shift in the *F2* domain (e.g., “hot”~”hat”) as a function of talker participation in NCCS. Finally, the two studies will come together in an argument of sociophonetic hearer-mediated sound change.

## **1.2 VARIATIONIST SOCIOPHONETICS**

In the 1960s and 70s sociolinguists began adopting the notion that there are sources of variability in language that can be traced to specific extralinguistic, social factors such as class, gender, race, and age. Following the results of his study of copula deletion in African American Vernacular English (AAVE), Labov (1969) suggested that these sources of variability can be quantified by means of the so-called “variable rule.” He annotated variable rules with probabilistic weights indicating how much a condition favors or disfavors the application of the rule. Subsequently, these conditions were expanded to include a wide range of social factors. Soon, Labov’s methodology provided the theoretical foundation for the branch of linguistics known as variationist sociolinguistics.

Since the early days of the discipline, many studies have focused on phonological variability (e.g., Labov (1972)) in phenomena such as deletion, epenthesis, devoicing, and other acoustic events perceivable by a trained phonetician. Later, researchers realized that in order to pursue this analysis to deeper levels, they had to study lower-level phonetic detail as well. While many sociolinguists still relied on impressionistic judgments of allophonic-level phonetic quality, a growing number of researchers employed acoustic analysis of speech, which enabled them to find patterns and relationships beyond those available with traditional methodology and instrumentation.

For example, Eckert (1999) explored the variability in the production of the second formant of /Λ/ as a marker of group identity. Bailey and Thomas (1998) provided a detailed acoustic account of AAE (African American English), while Preston, Evans, Ito,



and Jones (2000) explored the Northern Cities Chain Shift (NCCS) in Michigan from an acoustical point of view. As a result of this in-depth pursuit of acoustic detail, a new sub-discipline of sociolinguistics known as “sociophonetics” has emerged.

### **1.2.1 This dissertation as a work in variationist sociophonetics**

This dissertation is a work in sociophonetics. It aims at discovering, analyzing, and explaining low-level phonetic detail and its correlation with social factors such as sex, and region, both in terms of speech production and perception. It employs rigorous methodologies in the process of data acquisition and analysis. The acoustic features under investigation are most probably beyond the talkers’ linguistic awareness, which makes the data particularly interesting, as our understanding of the role of low-level phonetic detail brings us one step closer to understanding the mechanisms of language variation and change.

This dissertation is also a work of discovery. With few theory-dependent assumptions, it demonstrates that there exist certain, previously unexplored, sources of sociophonetic variability. It also shows that this variability is not random. On the contrary, it is systematic in ways consistent with milestone works in variationist sociolinguistics.

### **1.2.2 Current state of the field of sociophonetics**

#### **1.2.2.1 Speech production**

There is no doubt that sociophonetics has been dominated by empirical studies of speech production. Ranging from small studies of local variation (e.g., Anderson (2002)) to large projects such as the Telsur Project (Labov, Ash, & Boberg, 1997), there have been

numerous studies of vowel production, many of which used computer-assisted acoustic analysis. Sociolinguists have acquired a substantial knowledge of American English vowel systems, at least within the two-dimensional space delimited by the first and second formant (see 1.3.2.1). What still remains largely unknown, however, is how other articulatory-acoustic features, such as vowel nasalization fit in with the existing sociolinguistic research paradigm.

### **1.2.2.2 Speech perception**

Sociophonetic speech perception studies are still rare. One of the first rigorous and methodologically advanced studies in sociophonetic speech perception was a study by Graff, Labov, et al (1983). Using digitally manipulated speech samples (the onset of the diphthong /au/) they were able to obtain different judgments of vowel quality across ethnic lines. However, as Preston (1999) notes “such specific experimental procedures do not seem to be common in this research paradigm” (p. 7) that is at the interface of sociophonetics and speech science. As will become apparent later in this dissertation, this is not to be confused with perceptual dialectology, which has been very prolific, and which often provides inspiration for more phonetically focused, experimental studies.

## **1.3 SOCIOPHONETIC STUDIES OF SPEECH PRODUCTION**

### **1.3.1 Quantitative studies of consonantal variation**

#### **1.3.1.1 Impressionistic studies and feature dichotomy**

Sociophonetic studies of speech production attempt to find phonetic correlates of socially constructed linguistic behavior, and segmental features have been the most common object of acoustic scrutiny. Labov (1972) investigated the distribution of /r/ in data

collected in three New York department stores. This classic study of consonant deletion relied on impressionistic judgment of the presence or absence of /r/ in stressed syllables (e.g., “fourth,” “floor”). Labov examined a broad range of predictor variables, such as social class and gender, to establish their correlation with a particular phonetic (or allophonic) variant of /r/.

Many similar studies have been done since. In a typical case (e.g., Guy (1980)), the investigator collects pronunciations of target tokens as well as extralinguistic information about the speaker and the circumstances in which the utterance was made (metadata). The dependent variable is usually dichotomous (e.g., glottalization of stop consonants in word final positions) and its value depends on the investigator’s impressionistic judgment. Logistic regression analysis is often employed to explore possible relationships across the variables (Sankoff, Rand, Rousseau, Hindle, & Pintzuk, 2004).

#### **1.3.1.2 Limitations of impressionistic studies based on feature dichotomy**

The limitations of such a design are twofold. First, the value of the dependent variable is assigned by the researcher’s auditory impression of the token. While many phonological theories assume the binary nature of contrastive sounds (e.g., Chomsky & Halle (1968)), modern theories of speech production claim that the speech signal consists of continuously time-varying frequency components (e.g., Stevens (1972)). Human speech processing works on at least two levels – across category, sometimes referred to as contrastive or phonemic, and within category, also known as non-contrastive or allophonic. While listeners have generally no difficulty decoding the contrastive building blocks of speech, they can have great difficulty consciously evaluating variable sounds

within a single phonemic category (Lieberman & Mattingly, 1985). Thus, proper identification and classification of those low-level, articulatory and acoustic details, even by a trained phonetician, is prone to be inaccurate.

The other problem with this kind of design follows directly from the principles of speech production and perception described above. A number of phonetic features are, by nature, continuous. Voice onset time, formant transitions, release burst, intensity, or coarticulatory assimilation are best represented on an interval, rather than categorical scale (e.g., the presence vs. absence of a feature). While binary logistic regression can produce very robust results (Paolillo, 2001), the reduction of continuous criterion variables to a binary form may lead to imprecise prediction claims.

### **1.3.2 Quantitative studies of vocalic variation**

#### **1.3.2.1 The vowel quadrilateral as a metaphor for vowel space and movement**

Sociophonetic investigation of vowel production has produced impressive results. Since Labov Yeager, and Steiner (1972) introduced acoustic analysis to dialectal variation, sociolinguists have increasingly utilized instrumental techniques. Many American dialects (including NCCS) have been described in phonetic terms. The Atlas of North American English is one of the most influential and largest sociophonetic projects in modern dialectology. Labov and his colleagues (1997) collected speech samples by phone and analyzed them acoustically. The first and second formant of the steady state of each vowel and diphthong were recorded and plotted against a perceptual vowel **Q**uadrilateral, first proposed by Jones (1964) and later adopted by Ladefoged (1967) (Figure 2).

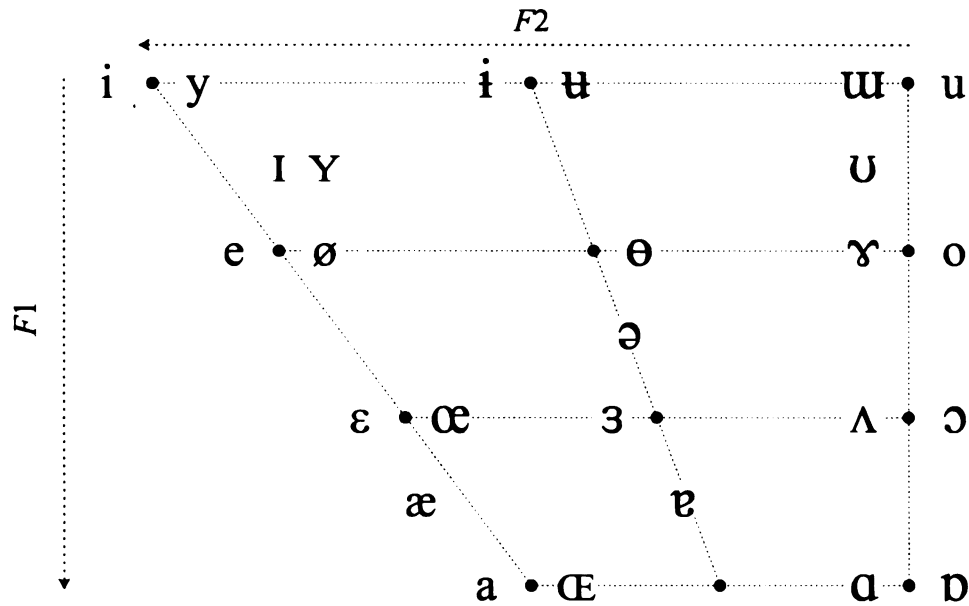


Figure 2. Vowel quadrilateral with IPA-style phonetic symbols

The  $F1/F2$  plot with reversed axes (values increase right-to-left and top-to-bottom – see the arrows labeled “ $F1$ ” and “ $F2$ ” in Figure 2) corresponds closely to the perceptual space proposed by Jones. For example, the high front tongue position for /i/ matches the low  $F1$  and high  $F2$  values of /i/. This representation of vowel space has given rise to an entire tradition of vocalic variation analysis in sociophonetics. The major premise of NCCS, for instance, is that vowels move in this space along the height ( $F1$ ) and frontness ( $F2$ ) dimensions. For instance, Preston and Evans (2000) created a degree-of-shift scale that they used to determine the degree to which talkers participated in NCCS. The scale was based on  $F1$  and  $F2$  differences in Hz, but spatial relations such as “higher than /ɛ/” were used to great explanatory effect.

### **1.3.2.2 Beyond *F1* and *F2***

The two-dimensional sociophonetic analysis of vowels has proved successful in establishing general dialectal boundaries (isoglosses) (e.g., Fridland (1998)). However, this type of analysis is not without limitations. For instance, it disregards a good deal of spectral information, such as formant bandwidths, *F0*, or the third formant frequency (Thomas, 2000). It also does not respond to current theories of speech perception, which emphasize the importance of critical band analysis in auditory processing (Chistovich, 1985), and contextual factors that may influence perceptual decision-making (Johnson, 1989). There is no doubt that much remains to be discovered by studying other acoustic phonetic features of vowels. As will be seen later in this work, aerodynamic studies hold a promise of interesting, new findings, as do quantificational studies of formant trajectories.

### **1.3.3 Sociophonetic accounts of suprasegmental features**

Segmental variation is not the only object of sociophonetic scrutiny. Perceptual and attitudinal studies, such as Preston (1996), have demonstrated that listeners often categorize dialects and sounds with descriptive terms such as “fast,” “slow,” “nasal,” “drawl,” or “twang,” which seems to extend folk linguistic perceptions well beyond phonetic segments. Sociolinguists often try to translate these intuitive labels into quantifiable suprasegmental features of voice production. However, while many listeners have little trouble identifying stereotypical dialectal differences based on intonation, pitch, rhythm, and phonation characteristics, these features still seem illusive to sociophoneticians (Thomas, 2000).

One of the most serious obstacles in sociophonetic research of suprasegmental variation is a general difficulty in quantifying suprasegmental events. Intonation, for instance, is variable within a broad range of frequencies over time. This variability is correlated with a number of linguistic and extralinguistic behaviors, such as semantics, style, register, or emotion. Despite the existence of notation systems, such as ToBE (Beckman & Hirschberg, 1994), quantificational sociophonetic work in suprasegmental features of speech production has not been nearly as extensive as the work on segments.

Several notable exceptions in contrastive dialectology include the work of Ee Ling, Grabe, and Nolan (2000) whose findings provide acoustic data (duration and  $F1/F2$  space characteristics) to support the phonologically hypothesized differences in rhythm patterning between British English and Singapore English.

## **1.4 SOCIOPHONETIC STUDIES OF SPEECH PERCEPTION**

### **1.4.1 Perceptual dialectology**

There has been a great deal of research in the area of perceptual, or folk dialectology (e.g., Preston (1996)). Perceptual dialectology is primarily concerned with common (folk) perceptions of language and language varieties. It is most interested in the relationship between language (or dialect) and various social constructs associated with it, such as identity, regionalism, standardness, gender, and ethnicity. In their attempt to study the nature of such beliefs and attitudes, researchers have often crossed over into other areas of scientific inquiry, such as social psychology (Williams, 1976), cultural geography (Preston, 1996), speech science (Purnell, Idsardi, & Baugh, 1999), and cognitive science (Nussbaum & Morin, 1993). The results of perceptual research have shown that

perceivers generally hold strong stereotypical views of dialects and their speakers, and that those views may have a considerable influence on speech perception.

### **1.4.2 Experimental studies of speech perception**

Sociophonetic studies of speech perception have been less common than both perceptual dialectology and production studies. Thomas (2002) attributes this to the lack of technological expertise that is necessary to design and carry out perceptual experiments. However, some of the recent results of perceptual research have encouraged sociophoneticians to venture into the world of speech science.

#### **1.4.2.1 Social information and speech perception**

Niedzielski (1999), in a sociophonetic speech perception study, found that judgments of vowel quality may be influenced by nationality labels. She presented 41 Detroit area residents with a set of “re-synthesized” stimuli.

The subjects were asked to choose from a set of re-synthesized variants the one that best matched the vowel heard in the tape-recorded speech sample of a fellow Detroiter. Half the respondents were told that the speaker they heard on the tape was from Detroit, and half were told that he was from Canada. In the first experiment involving the word “house”, and more specifically, the diphthong /au/, nationality labels significantly influenced the pattern matching task.

The remaining two experiments, involving the words “pop” and “last” did not obtain significant results in the same way. However, these two experiments demonstrated a



really interesting phenomenon. The word “pop” contains the vowel /a/, which is one of the vowels that are believed to be participating in the Northern Cities Chain Shift (NCCS) (Labov et al., 1972). Figure 3 shows the way in which the vowels are supposed to be moving within the two-dimensional space.

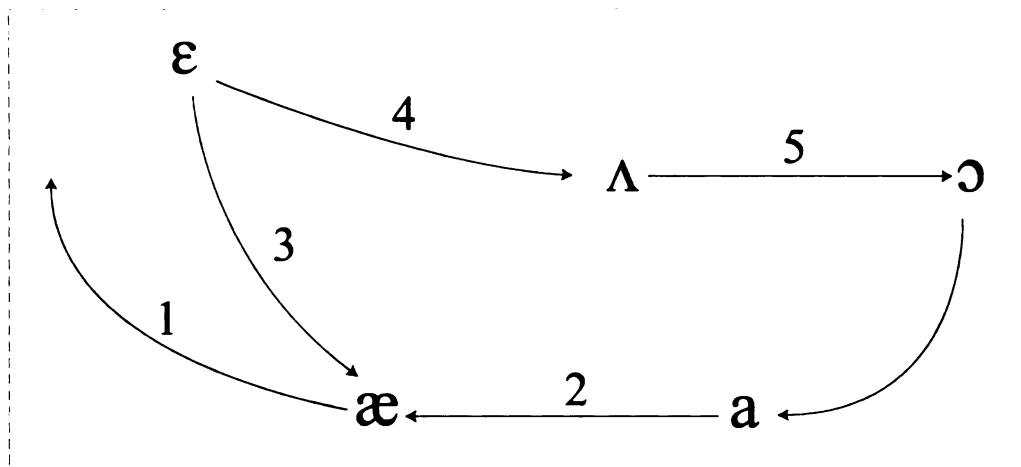


Figure 3. NCCS vowels and their movement within the two-dimensional  $F1/F2$  space

Detroit residents, no doubt, have been exposed to NCCS-influenced speech. It is likely that they themselves also sound this way. Yet, when presented with re-synthesized samples of the word “pop” they overwhelmingly choose a synthetic “canonical” version of /a/, in preference to a version that mirrored Detroiters’ productions. “Canonical” is the term Niedzielski used to describe vowels whose  $F1$  and  $F2$  were similar to those found by Peterson and Barney (1952). Figure 4 shows the response patterns.

Niedzielski’s account of this phenomenon is consistent with those found in perceptual dialectology, such as Preston (1999). Niedzielski argues that Detroiters consider themselves to be speakers of Standard English and, therefore, are reluctant to recognize an NCCS-influenced vowel as similar to their own. Instead, they select the canonical

form, demonstrating that their perception of a vowel variant within the same linguistic category (or phoneme) is sensitive to changes of 200 Hz along *F1* and *F2*.

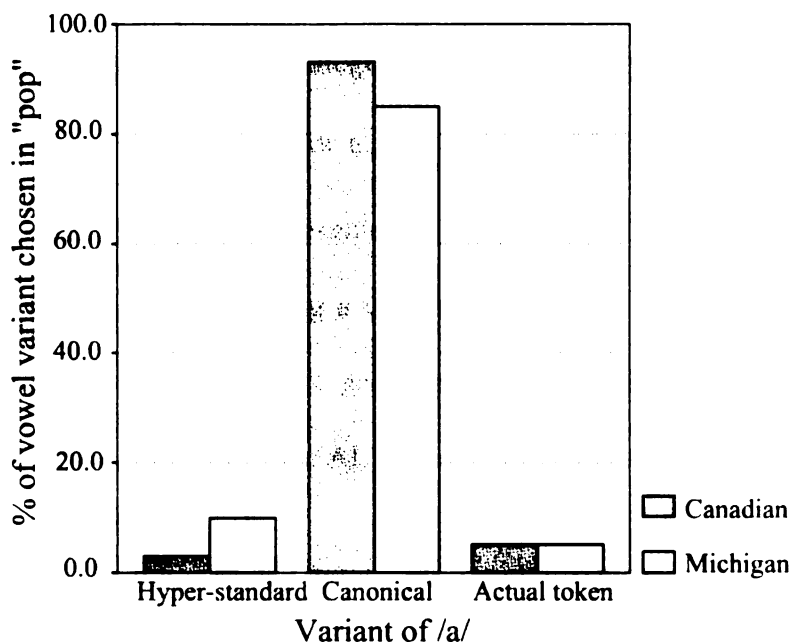


Figure 4. Responses to the word "pop" (From Niedzielski (1999) with permission)

#### 1.4.2.2 Vowel quality and vowel identity: “within category” and “across category” speech perception

The distinction between studying vowel quality and vowel identity is important to the understanding of this dissertation. Niedzielski (1999) studied vowel quality. Each time the respondent was presented with a stimulus, it was known which word (and hence which vowel) would occur. The listeners’ choice was about a preferred token (a variant) of the vowel type.

Alternatively, acoustic cues might be varied over a range that causes a perceived change in vowel identity, that is a perceptual shift from one phonemic (contrastive) category to another (Stevens, 1985).

### 1.4.2.3 Vowel quality and dialectal stereotype

It is no secret that listeners have a certain sense of regional dialect. However, what still remains illusive is which specific pronunciation features contribute to regional dialect identification. Within category vowel variability is a mainstay of sociophonetics, and the pronunciation of /aɪ/, for example, is one of the more interesting such features. Plichta and Preston (2004) presented 96 respondents from different parts of the United States with a task of matching a variant of the word “guide,” spoken by a male and a female talker, with one of nine geographical locations along the North-South dimension from Saginaw, MI to Dothan, AL (the stimuli were randomized and repeated four times each).

They used the Linear Predictive Coding (LPC) analysis/re-synthesis method (see also 3.2.5.1) to produce a 7-step continuum of the word “guide” by simultaneously varying the first formant (“raising”) by a total of 150 Hz and the second formant (“fronting”) by a total of 550 Hz (see also 3.2.5.2). The female and male pronunciations remained parallel in how their formant trajectories changed throughout these continua. Figure 5 shows formant tracks of  $F1$  and  $F2$  at each of the seven steps. The dots of each shade of gray represent formant trajectories at each of the seven steps of monophthongization and fronting.

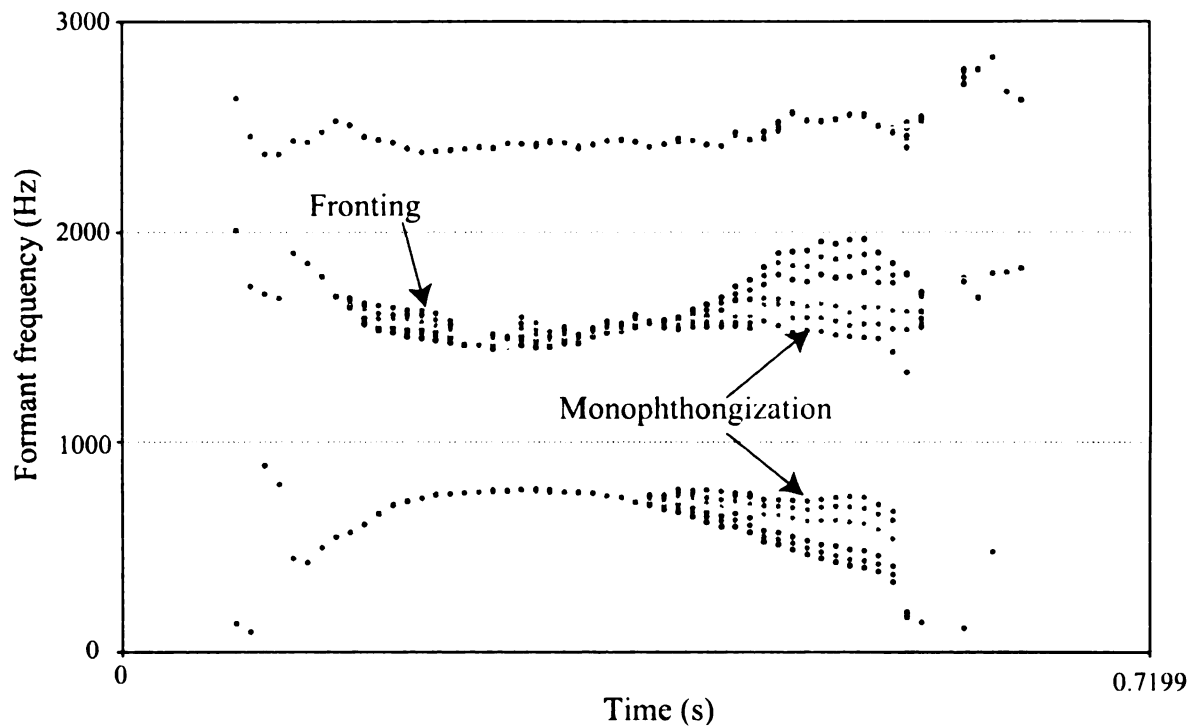


Figure 5. Formant tracks of the 7 variants of the male pronunciation of the word "guide"

This study is similar to that of Niedzielski (1999) in that it also focuses on “within category” perception. Despite the substantial overall changes in *F1* and *F2* frequencies between the first and the last step, the sound /aɪ/ retained its contrastive (phonemic) identity. However, despite relatively small differences in *F1* and *F2* frequencies between the neighboring steps of the continuum (less than 30 Hz along *F1* and 60 Hz along *F2*), listeners responded differently to each stimulus. The responses changed over the series in a nearly continuous manner, as if mapping the acoustic continuum onto the geographical one. Figure 6 shows mean responses to the male and female voices. Each stet (except step 1 and 2) is significantly different from its neighbor.

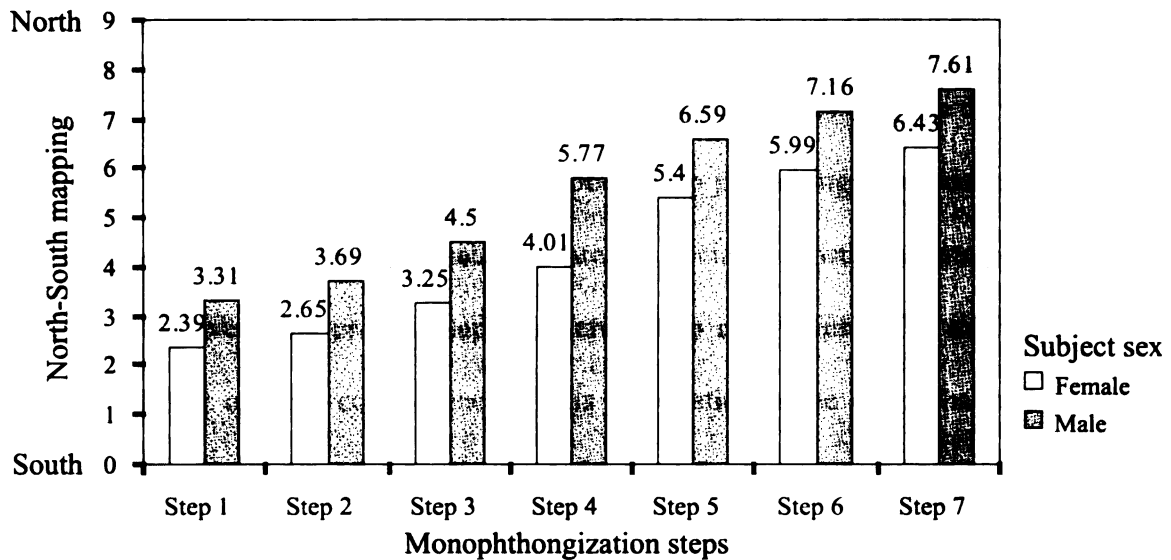


Figure 6. Mean responses to male and female voices

As evident in Figure 6, the female voice (white bars) was not rated as “Southern” as the male voice (gray bars). This, along with the overall sound-region mapping result, contributes to the general conclusion that, in experimental conditions, listeners are capable of distinguishing among very small absolute changes in the frequency domain of the first and second formant, even “within category.” Moreover, this perception is systematic and consistent with the common stereotypes of Southern and Northern speech. Niedzielski’s study demonstrated that changes of 200Hz in the first and second formant frequencies carry dialectally salient weight; this study showed that even smaller increments in formant frequencies play a role in regional dialect identification.

#### 1.4.2.4 Gender information and speech perception

Strand and Johnson (1996) were interested in “across category” perceptual effects, that is in identity shifts. They studied the effects of gender on speech perception and demonstrated that perceivers produced differing consonant identity responses to the same

stimuli (/s/ and /ʃ/) depending on the perceived gender of the talker. This can be explained by the fact that men have lower turbulence frequency in these consonants than women. The authors suggest that the bottom-up processing of acoustic information interacts with higher-level information related to socially constructed stereotypes about gender. Moreover, this process happens across two modalities – the auditory and the visual. What is of most consequence to this dissertation is the claim that the perceived information about the talker's gender, along with the pre-conceived, and implicit knowledge of men sounding different than women, appears to be tightly integrated in the speech perception process.

#### **1.4.3 Needed research in perceptual sociophonetics**

Though interest in the sociolinguistic content of speech signals appears to be growing, sociophonetic perception studies are still rare. Perhaps the most intriguing questions that perception studies might answer are (1) how much sociolinguistic information is conveyed in the speech signal (bottom-up information), and (2) to what extent external knowledge and expectations (top-down information) influence speech perception. Better understanding of such processes can help solve the age-old mystery of sociolinguistic salience. In addition, sociophonetic research in speech perception can play a crucial role in broadening our understanding of role of sociolinguistic information in the talker

normalization process.<sup>1</sup> Thomas (2002) points out that speech perception is likely to continue having an increasingly important role in sociophonetic research.

#### **1.4.4 Speech Science –similar questions, different methodologies**

##### **1.4.4.1 Introduction to perceptual research in speech science**

Investigations of speech production and perception have been of interest to speech scientists for some time. In the 1950s, when the first speech synthesizers and spectrographic speech analyzers became available, speech scientists obtained a great deal of new insight into the details of speech production and perception. Several theories of speech perception have evolved based on this work. These include the Motor Theory (Liberman & Mattingly, 1985), Quantal Theory (Stevens, 1989), and the Fuzzy Logical Perception Model (Massaro, 1998). Just as sociolinguists recognize the need to venture into speech science, speech scientists acknowledge the importance of integrating the sociolinguistic dimension in their work (Nussbaum & Morin, 1993).

This interest was first sparked by Ladefoged and Broadbent (1957)<sup>2</sup> who demonstrated that the final output of the perception process is contingent upon the integration of sensory input with a (more abstract) vocalic frame of reference. Details of this study are summarized in 3.1.1.5. More recently, Purnell and Kopplin (2003) investigated the role of source and filter characteristics in the identification of talker ethnicity. They found that filter characteristics (formant values) play a crucial role in distinguishing between General American English and African American English.

---

<sup>1</sup> See discussion of the Ladefoged and Broadbent (Ladefoged & Broadbent, 1957) study in CHAPTER 3.

<sup>2</sup> See 3.1 for a more detailed discussion.

#### **1.4.4.2 Dynamic constancy**

An influential theory of speech perception argues that discrete units of speech, often referred to as phonemes, are encoded by the talker into the highly variable speech signal. These units are then decoded by the listener, effectively correcting for the articulatory/acoustic variability and separating linguistic elements from noise (Lieberman & Mattingly, 1985).

This phenomenon, often referred to as “dynamic constancy,” bears close resemblance to the variable rule theory of variationist sociolinguists. One might even think of those two fields as complementary since one deals with physiological and the other with social correlates of human speech. However, the seemingly simple marriage of those two ideas has not yet been widely acknowledged.

#### **1.4.4.3 Categorical speech perception**

As mentioned in 1.4.2.2, speech perception studies can investigate “within category” or “across category” effects. Speech scientists have been most preoccupied with “across category” effects. In the 1950s, perceptual research revealed that the perception of speech sounds is not linear and that it differs greatly from the perception of other sounds in the time/frequency domain.<sup>3</sup> Specifically, continuous acoustical changes in the auditory input can result in sharply discontinuous, categorical perceptual shifts. There have been many categorical perception studies, but there is one that is of most importance to this dissertation (also see 3.2.7 later in CHAPTER 3).

---

<sup>3</sup> Such non-speech sounds are sometimes referred to as “non-speech analogs.” The perception of “within category” human sounds have been very similar to that of non-speech analogs.



Pisoni (1973) performed a series of perceptual tests with vowels of varying length (300 ms and 50 ms) and with consonants of varying voice onset time (/p~/b/). He discovered a very categorical perception of consonantal contrast and a slightly less categorical perception of vowels. However, as vowels became longer, their perception patterns became similar to those of consonants. This was also the case when there was an extra stimulus presented between subsequent vowel tokens. Figure 7 shows psychometric functions from the mean aggregate responses to long vowels and VOT variability in /p~/b/ pairs. The vowel identification task (left) produced results similar to those of voicing identification. Even though the stimuli increase incrementally and steadily in equal physical steps, the change in auditory response occurs rapidly and discontinuously - when the stimulus reaches the area close to the category boundary of /i~/l/ and /p~/b/, respectively. The functions of the /p~/b/ contrast appear to have slightly steeper slopes, which indicates a more categorical perceptual process (see also 3.2.7.1).

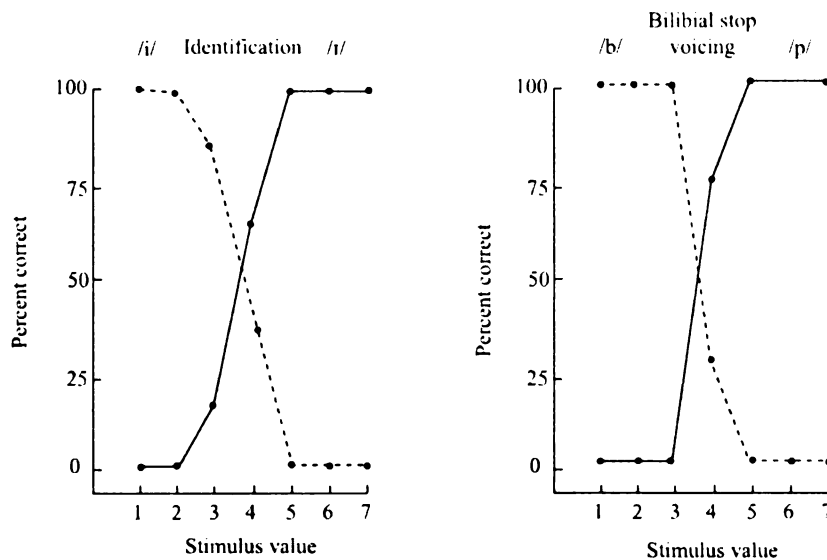


Figure 7. /i/ and /l/ and /p/ and /b/ identification and discrimination results (From Pisoni (1973) with permission)

#### 1.4.4.4 Integration of acoustic cues in speech perception

In a classic stimulus integration study, Fitch, et al (1980) discovered that the perception of category (or phoneme) boundaries shifts as a function of the consonantal environment in which they occur. Two “slit”~“split” continua were synthesized with varying amounts of silence between the high frequency noise of /s/ and the onset of the voiced remainder of the words (/lit/), as can be seen in Figure 8. The results of the experiment showed that the boundary between “split” and “slit” changed depending on the length of the silence interval. The longer the interval (right), the more likely were the respondents to identify the stimulus as “split,” and vice-versa. This change of category boundary (or phonemic boundary) depending on external factors (in this case the duration of the silence interval) was termed “phonetic trading relations.”

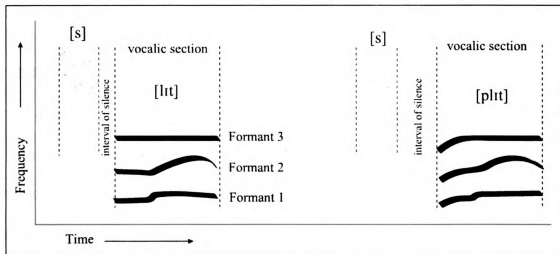


Figure 8. Simplified spectrograms of the stimuli used in the experiment From Fitch, et al (1980)

## **1.5 ORGANIZATION**

The present dissertation applies technological and theoretical advances from speech science to the study of sociophonetics – specifically to the study of NCCS. It includes two studies, one is speech production, and the other is speech perception.

### **1.5.1 The speech production study**

The speech production study presented in CHAPTER 2 discusses some problems with the methodologies and instrumentation of previous vowel analyses and presents a new method of sociophonetic analysis of NCCS-influenced vowels. A new research methodology is also proposed. It is based on aerodynamic measurements of vowel nasalization and it investigates the distribution of vowel nasalization as a sociolinguistic marker. This method is meant to provide a tool, complementary to current approaches, to further our understanding of this intriguing vowel shift.

### **1.5.2 The speech perception study**

The speech perception study in presented in CHAPTER 3, focuses on “across category” perceptual influences of dialectal information conveyed by the speech signal. It demonstrates that the fronting of /a/, one of the markers of NCCS (Labov et al., 1972), bears a significant amount of perceptual weight among NCCS speakers, which, in turn, provides experimental validity to previous studies in perceptual dialectology, such as Labov (1991), which reported a perceptual confusion between /a/ and /æ/ between NCCS and non-NCCS dialect speakers.

### **1.5.3 Comments and conclusions**

The last chapter brings the two studies together into a unified account of hearer-mediated sound change. It will be argued that the re-shaping of the NCCS vowel space is correlated with vowel nasalization and that the movements of /æ/ and /ɛ/, in particular, result from an on-going speaker-hearer negotiation of vowel contrast influenced by nasalization. Similarly, the salience of /a/-fronting is negotiated via the hearer's speech community. One's speech community must be perceptually ready before it can reproduce and spread new pronunciations of /a/.

## **CHAPTER 2**

### **VOWEL NASALIZATION AND NCCS**

#### **2.1 INTRODUCTION**

This chapter presents an analysis of sociophonetic variation in Michigan. The analysis goes beyond the two-dimensional formant frequency measurements associated with the vowel quadrilateral. The specific goal of this chapter is to demonstrate, with quantitative analyses, that new insight into sociophonetic variation can be gained from aerodynamic measurements of oral and nasal airflow.

In particular, the study will show that the output of traditional computer-assisted sociophonetic analysis of vowels is subject to significant variability depending on data acquisition techniques used, and that this variability is particularly high in the LPC measurements associated with *F*1. The problem becomes even more serious when applying standard LPC analysis to vowels with complex spectra, such as nasalized /æ̃/. Nasalized /æ̃/ is particularly difficult because a low-frequency nasal formant, typically, comes below the first oral formant, thus often forcing the LPC filter into returning a low-frequency value misinterpreted as the first oral formant of /æ/. If this technique is applied to a large speech corpus, subsequent statistical analysis may lead to a significant biasing of the data and to a misclassification of /æ/ as “raised” (marked by lower *F*1 frequency) when /æ/ is in fact nasalized.

The study will further show that vowels pronounced by many Michiganders have rather complex spectral features, and that they often show the presence of nasal peaks and anti-

formants. Moreover, these nasal features occur in nasal, as well as non-nasal environments and are distributed along sociophonetic isoglosses of region and sex. Finally, quantificational evidence will be provided to support the claim that the amount of nasalization in non-nasal environments (e.g., “dad”) is correlated with talker participation in NCCS.

## **2.2 FRONTING AND RAISING**

As mentioned in 1.4.2.1, NCCS is defined as a movement of vowels in the two-dimensional space delimited by *F1* and *F2* (see Figure 3 in 1.4.2.1). This has given rise to the nomenclature of physical movement, which is often used in sociophonetics (Preston et al., 2000). For example, if a vowel is said to “front,” it is to be interpreted that the *F2* of this vowel increases. Similarly, if a vowel is said to “raise,” it is because its *F1* decreases. Figure 9 provides a visualization of this relationship. The vowel /æ/ is “raising,” i.e., its *F1* is decreasing, while the vowel /a/ is “fronting,” i.e., its *F2* is increasing. Note that the exact path of *F1* in Figure 8 is not perfectly vertical, as /æ/-raising in NCCS is often believed to have a fronting component, as well (Gordon, 1997).

This framework assumes that there once was a different vowel system in Lower Michigan and that this vowel system is now in flux, whereby vowels are “moving,” occupying new spots and stimulating further movement of other vowels in the vowel quadrilateral. Moreover, these movements are conjectured to be happening in a specific order, the vowel /a/ being the first to move, though this particular process has been disputed by diachronic research (Stockwell & Minkova, 1997).

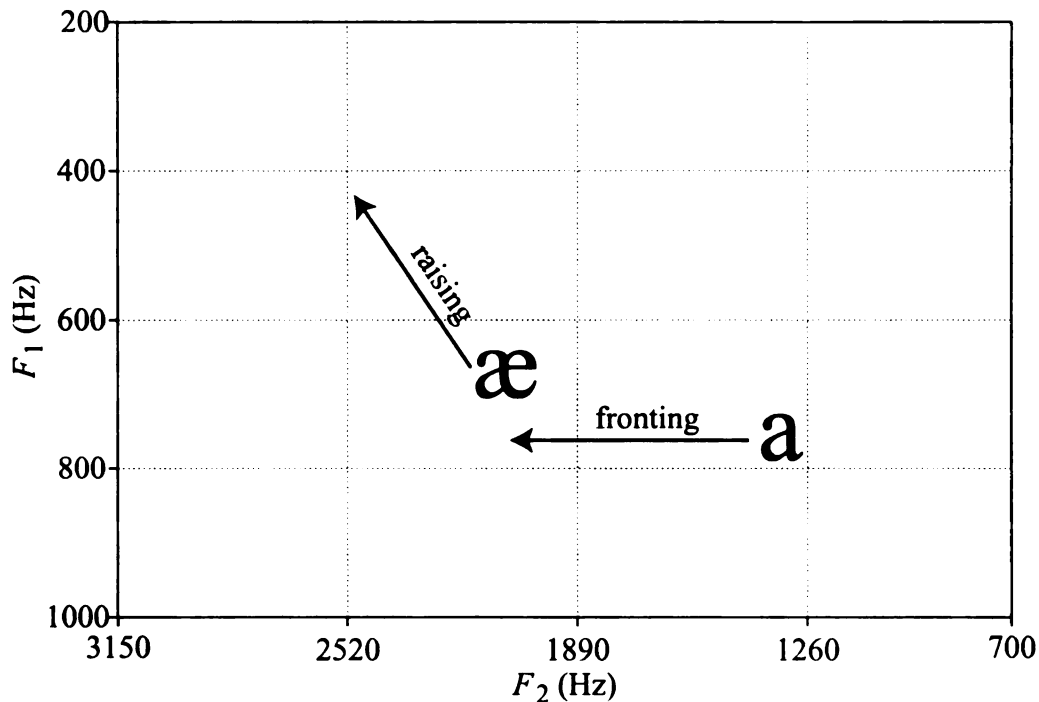


Figure 9. The fronting of /a/ and the raising of /æ/

## 2.3 PROBLEMS WITH TRADITIONAL ACOUSTIC ANALYSIS OF VOWELS

### 2.3.1 Data acquisition problems

It is conjectured here that data acquisition methods may have significantly influenced past acoustic analyses of NCCS. An experiment was conducted to see how three different field recording methods would influence acoustic analysis. The first two methods are very common in sociophonetic literature: (1) signal acquisition with a Marantz (or comparable) cassette recorder with a built-in microphone; or (2) recording speech with a MiniDisc recorder and an omni-directional lavalier microphone. The third method, still rather uncommon, involves (3) recording speech with a head-set, flat response microphone and a 24-bit digital recorder.

### **2.3.2 Data acquisition with a standard analog Marantz recorder**

Marantz portable cassette recorders have been used in the field for some time. They have a reputation for being sturdy and for producing high quality recordings, particularly for the purposes of news gathering and reporting. The recorder comes with a built-in condenser microphone, capable of capturing broadcast quality sound. Note, however, that audio quality understood in audiophile terms (e.g., “broadcast quality”) does not directly correlate with audio quality in acoustic-phonetic terms, as the former relies on subjective assessment of abstract sound properties, such as “clarity,” “brightness,” or “presence,” (Barlett & Barlett, 1998) while the latter refers strictly to the acoustic fidelity and detail of acquired sounds.

The potential problems with the Marantz-type recorder are three-fold. First, the omnidirectional microphone is too close to the recorder’s motor and tape transport mechanism. It therefore “picks up” a lot of low frequency noise. Figure 10 shows an LPC spectrum of the vowel /i/ superimposed on a narrow-band spectrum of the noise produced by the recorder.<sup>4</sup> It is evident that the spectral band of the first formant (labeled “F1”) is similar in level to the low-frequency components of the noise (labeled “Noise”). This may negatively influence LPC-based formant extraction.

---

<sup>4</sup> Other low-frequency noise generated by a computer fan or an air conditioner would have a very similar spectrum.



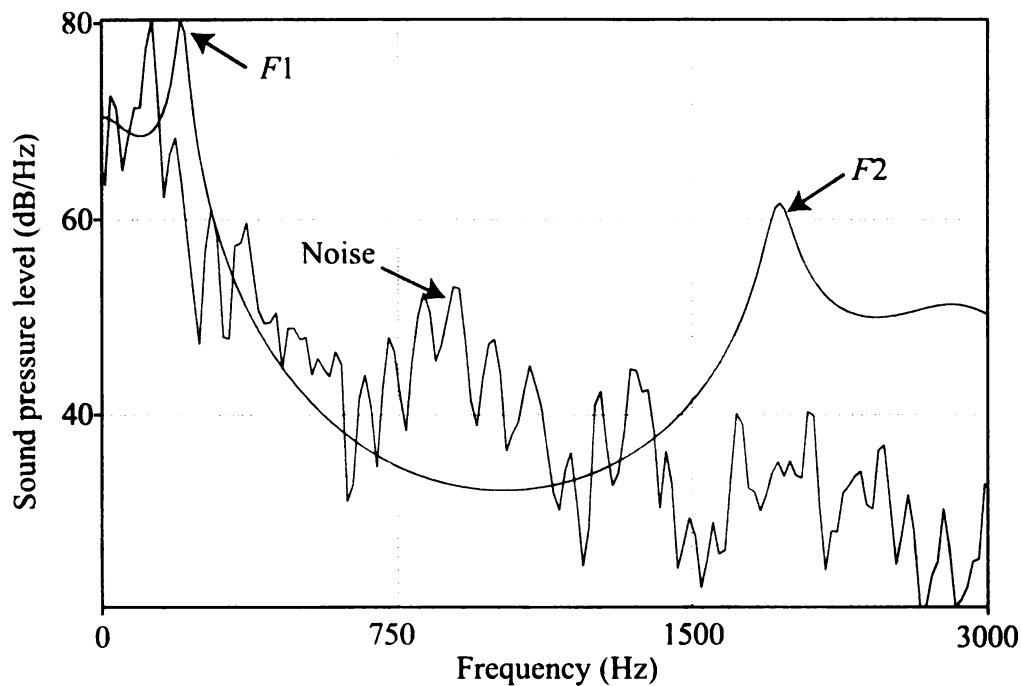


Figure 10. LPC of the vowel /i/ superimposed on the noise spectrum of the Marantz recorder

Second, the Marantz microphone, due to its omni-directional polar pattern, records noise from the environment, as well as the intended speech signal. If the interview is conducted near an air conditioner, open window, refrigerator, or any other source of low frequency energy, the amount of noise recorded on tape can jeopardize the reliability of acoustic analysis. Finally, this kind of microphone cannot be placed close to the talker's lips, which causes a great deal of attenuation (decrease in amplitude) of the speech signal before it reaches the microphone.

#### 2.3.2.1 60 Hz hum

Sixty Hz hum from power circuits is another common source of noise in field recordings, particularly when unbalanced microphone cables and non-grounded power cables are used. Figure 11 shows two spectra – one of the 60 Hz hum (solid line) and one of the

vowel /a/ in “job” (dashed line). The hum interferes with the speech signal in ways similar to those illustrated in Figure 10. This problem is not unique to Marantz recorders, though the popular PMD 101 model does not have a balanced microphone input and its power supply is, typically, not grounded, which makes this recorder particularly susceptible to 60 Hz hum. Also, often, in respondents’ kitchens and living rooms, grounded power outlets are not available, and the existing wiring used for refrigerators, TVs, air conditioners, and so forth, can potentially cause a great deal of interference. Therefore, the use of professional-grade, balanced, properly grounded, “XLR” (also known as “Cannon”) microphone interfaces and cables is necessary to ensure the least amount of electrical interference.

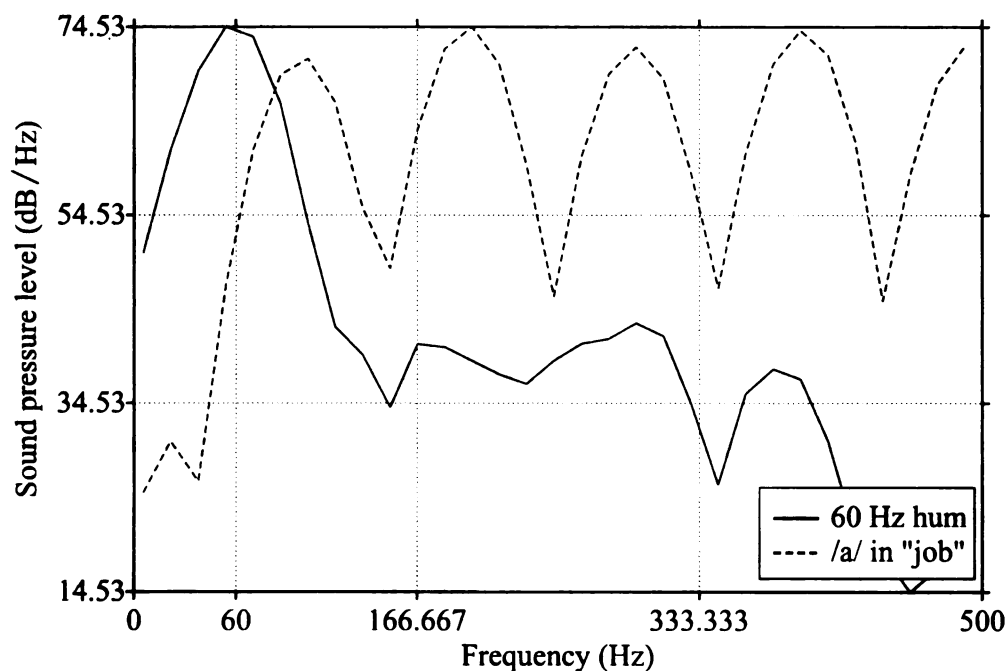


Figure 11. Spectrum of 60 Hz hum and the vowel /a/ in “job”

### **2.3.3 Digital recording with a MiniDisc player**

The Sony MiniDisc form factor first appeared on the market in Japan in 1992. Soon afterwards, due to an effective 1996 advertising campaign entitled “Where the Music Takes You“, the MiniDisc became the most popular portable consumer digital music system (Hunt, 1996). It was also embraced by the fieldworker community, despite the fact that it was never designed to be a quality recording device. Sony designed the MiniDisc recorder as an inexpensive digital option to its Walkman series of cassette players. The MiniDisc has at least three serious problems.

First, the speech signal is altered by the recorder (compressed) in ways over which the researcher has no control. MiniDisc recorders use a lossy psychoacoustic data compression format called ATRAC (Adaptive TRansform Acoustic Coding). Soon after the recorder captures and quantizes an acoustic signal, it converts it to the proprietary ATRAC format at the bit rate of 292 kbps (approximately 1/5 of the uncompressed PCM “CD quality” of 1.41 Mbps). All of the acoustic field data are processed by the algorithm, which is based on psychoacoustic principles whereby the signal is divided into three frequency sub-bands with different data reduction schemes involved in each of them (Pohlmann, 2000). This indicates non-linear digital signal processing whereby the original speech signal is altered in ways determined but by the best compression (compromise between “quality” and bit-rate) options dynamically selected by the algorithm.

Second, the standard MiniDisc recorder does not have a professional microphone interface (see 2.3.2.1 for a discussion on the importance of the balanced XLR microphone

interface). As a result, only a small number of amateur quality microphones can be used with it. It is equipped with an electret-condenser interface with so-called “plug-in power.” Some Sony microphones are compatible with it, as are a few of other brands (e.g., Audio-technica AT803b and AT822). While it is theoretically possible to connect a professional-grade microphone to the MiniDisc player, this is quite difficult, as it requires a special impedance matching in-line transformer.

Finally, an omni-directional, low-quality lavalier microphone that is most typically used with MiniDisc players, produces a noisy recording. Figure 12 shows a waterfall spectrogram of the word “hat” recorded with a MiniDisc recorder and an Audio-technica AT803b microphone. The spectrum appears to contain a great deal of extraneous noise and the formant peaks have unusually low intensity and wide bandwidths when recorded with MiniDisc hardware. MiniDisc recordings are digitized at a dynamically variable bit rate, but they can be converted to a PCM format at 44,100 Hz. The spectrogram in Figure 12, therefore, does contain acoustic information well over 3,000 Hz. However, this information contains very little spectral detail of interest to phoneticians.

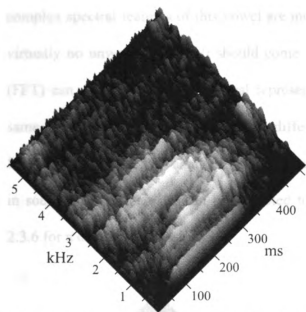


Figure 12 Waterfall plot of the word "hat" by an NCCS-influenced female talker

### 2.3.4 24-bit digital recording with a close-talking microphone

The third scenario involves a high quality, 24-bit, 48,000 Hz digital hard disc recorder with a flat-response, head-set microphone (Sennheiser HMD25-1). The Sennheiser microphone has been specially designed for recording speech in noisy conditions. It has a directional polar pattern, but despite being a close-talking microphone, it retains a flat response throughout the entire frequency range of up to 16,000 Hz. By being close to the talker's lips, it records speech with a more favorable signal-to-noise (S/N) ratio. The digital recorder, which connects directly to a laptop computer via the USB bus, is equipped with two professional-grade microphone inputs. It reproduces pure 24-bit sound, which contains more acoustic detail and less quantization noise than signals captured with 16-bit digital recorders (such as Tascam DA-P1 portable DAT recorder).

Figure 13 shows a waterfall spectrogram of the word "hat" recorded by the same speaker (see also similar plots in Figure 28), at the same time as the sample in Figure 12. The

complex spectral features of this vowel are more visibly strong and well defined; there is virtually no unwanted noise. It should come as no surprise that Fast Fourier Transform (FFT) can produce different spectral representations of what are all productions of the same vowel captured with these three different recording methods. Similarly, Linear Predictive Coding (LPC) analysis, which is the standard way of extracting formant values in sociophonetic analysis, can be expected to output significantly different values (see 2.3.6 for a detailed discussion).

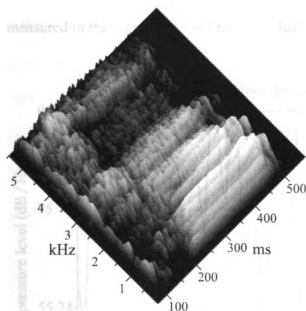


Figure 13 Waterfall plot of the word "hat" by an NCCS-influenced female talker

### 2.3.5 The role of the microphone

As mentioned in 2.3.2, 2.3.3, and 2.3.4, the role of the microphone in acquiring reliable speech signals is at least as important as the role of the recording device and the recording medium. The ideal microphone must have a wide and flat frequency response, and must not cause an increase in low-frequency amplitudes when placed close to the talker's lips (so-called "proximity effect"). Proximity effect is a natural consequence of

placing a dynamic, cardioid microphone close to the sound source. Interestingly, most manufacturers are not interested in eliminating proximity effect as the extra low-frequency boost is often desired by stage performers and newscasters (Huber & Williams, 1998). Figure 14 shows a spectrum of the vowel /a/ in “job” recorded with the AKG C-420 microphone placed approximately 6 cm from the talker’s lips. This sample shows a significant increase in amplitude in low frequencies around 100 Hz due to proximity effect. While the measurements of  $F1$  and  $F2$  in this particular case might not be significantly affected by proximity effect, the measurements of  $F0$  and other parameters measured in the vicinity of the first 2 or 3 harmonics will be considerably biased.

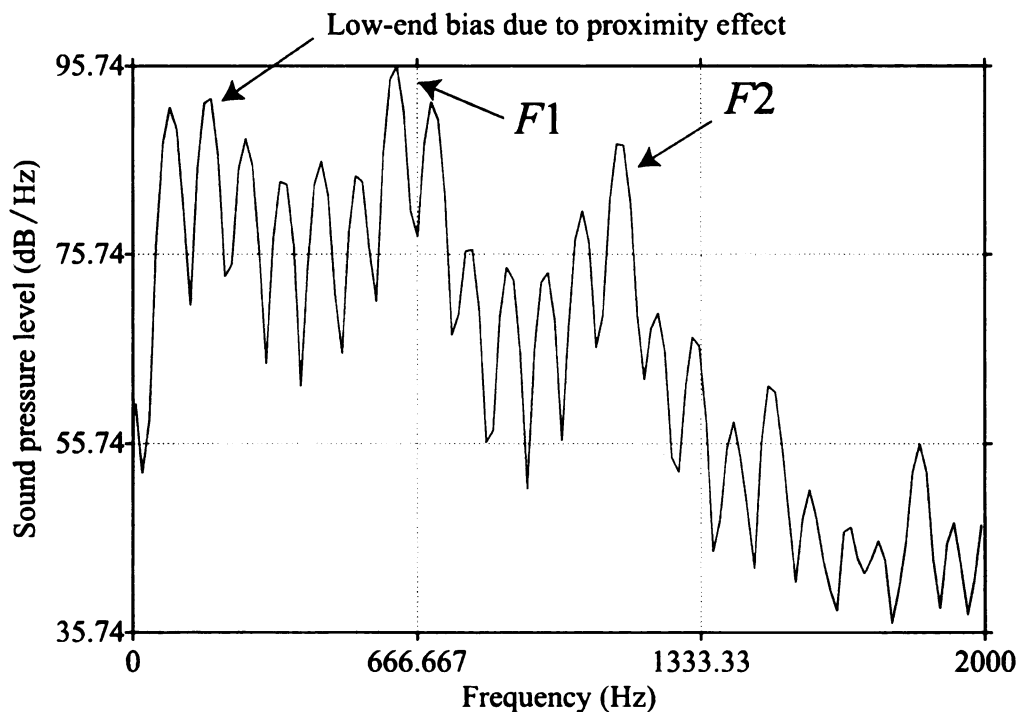


Figure 14. Low-end bias due to proximity effect of the AKG C-420 microphone

There are very few, specially designed microphones that retain broad and flat frequency response from 20 to 20,000 Hz. Such microphones, particularly when coupled with 24-bit

digital recorders, are capable of capturing high quality, unbiased speech signals. Figure 15 compares frequency response of Shure Beta 87a<sup>5</sup> and Earthworks M30 microphones, according to the manufacturers. The frequency response of the Shure Beta 87a microphone has two significant peaks – one around 5,000 Hz (solid line), and one around 100 Hz (dashed line). The increase in low frequency amplitudes (dashed line) occurs when the microphone is placed close to the sound source (below 6 cm). Conversely, the low-frequency amplitude decreases when the microphone is moved away from the sound source (above 60 cm). The Earthworks M30 microphone, on the other hand, retains a relatively flat frequency response throughout the entire human hearing range (dotted line), regardless of its distance to the sound source – subject to the limitations of the hardware used in the remainder of the recording circuit. Section 2.5.3.3 contains a detailed discussion on how proximity effect influences the analysis of low-frequency harmonics.

---

<sup>5</sup> Shure SM48, a dynamic microphone similar to Shure Beta 87a, was once distributed with the Kay Elemetrics Computerized Speech Lab (KayElemetrics, 1998).



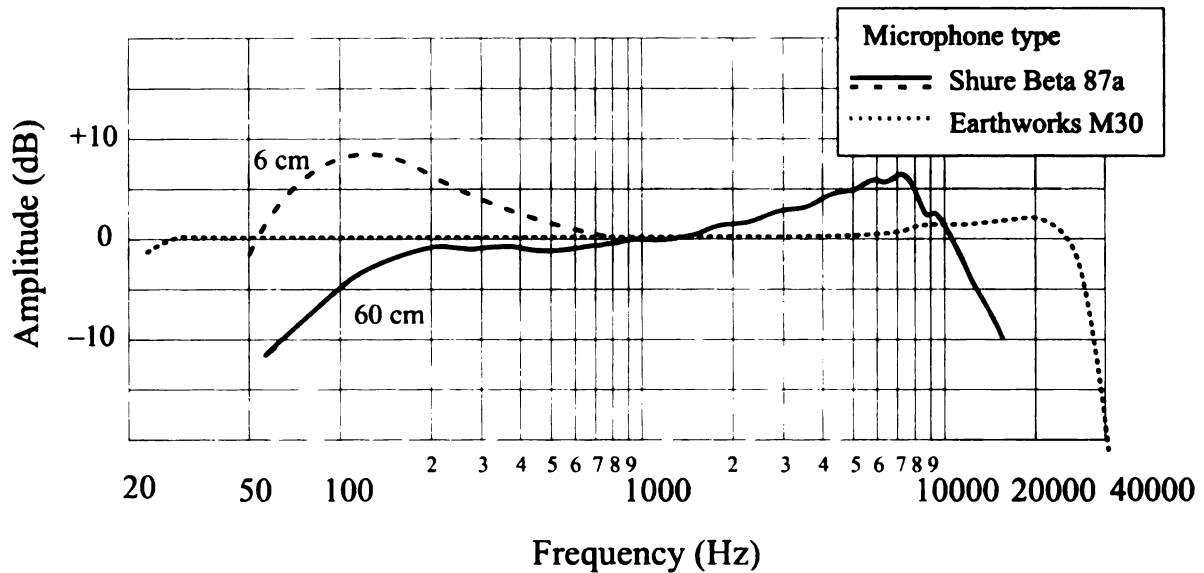


Figure 15. Frequency response of Shure Beta 87a and Earthworks M30

### 2.3.6 Different signal acquisition methods return different formant values

Given the information in the previous section, one might suspect that LPC analysis would output significantly different formant values across the three field recording methods. An experiment was designed to test this hypothesis. A female talker with an NCCS-influenced vowel system was recorded reading a wordlist containing words with four American English vowels (/a/, /æ/, /ɛ/, and /aɪ/). The first three vowels are participants in NCCS, and the fourth one, /aɪ/, was selected because of its diphthongal quality. The talker was recorded with a Marantz recorder, an Audio-technica AT803b lavalier microphone connected to a Sony MiniDisc recorder, and a Sennheiser head-set microphone plugged into a Sound Devices USB Pre unit simultaneously.

A total of 99 vowel tokens were recorded and transferred to a computer workstation. The first two sets were transferred via the digital S/PDIF<sup>6</sup> interface, and the analog recording was digitized at 24-bit and 48,000 Hz. The recordings were then downsampled to 16,000 Hz and analyzed acoustically. Acoustic analysis was performed by means of *Akustyk* (Plichta, 2004).

#### **2.3.6.1 How to avoid researcher bias in formant analysis?**

It is difficult to run a test comparing three different sets of recordings without researcher bias. This bias is caused by the subjectivity in identifying the steady state, an area of relative stability in the frequency domain, where formants are to be measured (see, for example Hillenbrand, Getty et al. (1995) for a more detailed discussion on steady state identification). Therefore, instead of identifying steady state and measuring formants statically, a dynamic approach was designed. The vowel nucleus was divided into 0.025s Gaussian analysis windows. Formants and bandwidths were measured and recorded for each such window. Based on these data, the software calculated mean formant and bandwidth values for the entire duration of the vowel, as well as their average cumulative variation based on the algorithm given below, where  $v$  is the average variation over time,  $x_n$  is the frequency at point  $n$  and  $d$  is vowel duration:

---

<sup>6</sup> S/PDIF (Sony/Philips Digital Interface) is a widely used digital audio interface capable of carrying stereo data quantized at 24-bit and 96,000 Hz.

$$v = \frac{\sum_{n=1}^T (x_{n+1} - x_n)}{d} \quad (1)$$

### 2.3.6.2 Statistical analysis of formants, bandwidths, and cumulative variation

Multivariate analysis of variance (MANOVA) was performed to test the hypothesis that the three recording modes return different formant values. The dependent variables included formant values in Hertz ( $F1$  through  $F3$ ), formant bandwidths in Hertz ( $B1$  through  $B3$ ), as well as total cumulative variation the first three formants in Hertz/s ( $F1v$  through  $F3v$ ), while the three different recording modes constituted the independent variables. MANOVA was significant for recording type,  $F(18,188)=8.318$ ,  $p<.001$ , Wilks'  $\Lambda=0.312$ . Because the overall MANOVA was significant, a series of post-hoc tests was performed to discover which acoustic parameters were significantly affected by recording type, and it was found that recording type was significant for all of  $F1$  parameters. The post hoc Bonferroni multiple comparisons test (at the significance level of  $p<0.05$ ) showed that all  $F1$ -related parameters were significantly different from one another across the three different recordings types. In addition, the tests found that all vowels individually, exhibited a similar kind of variability. Table 1 summarizes the results, where  $F1$  is the mean first formant frequency,  $B1$  is the mean first formant bandwidth, and  $F1v$  is the mean cumulative  $F1$  variation.

	Tests of between-subjects effects		
	<i>F</i> 1	<i>B</i> 1	<i>F</i> 1v
Overall	$F(2,102)=17.741$ , $p<.001$	$F(2,102)=17.741$ , $p<.001$	$F(2,102)=3.781$ , $p<.025$
/ɛ/	$F(2,21)=1.882$ , $p=.177$ (ns)	$F(2,21)=8.042$ , $p<.001$	$F(2,21)=.900$ , $p=.421$ (ns)
/a/	$F(2,21)=7.94$ , $p<.025$	$F(2,21)=4.544$ , $p<.05^7$	$F(2,21)=1.076$ , $p=.359$ (ns)
/æ/	$F(2,36)=6.12$ , $p<.025$	$F(2,36)=3.773$ , $p<.05$	$F(2,36)=4.15$ , $p<.025$
/aɪ/	$F(2,15)=8.737$ , $p<.025$	$F(2,15)=31.106$ , $p<.001$	$F(2,15)=2.972$ , $p=0.82$ (ns)

Table 1. Summary of statistical results of the data acquisition test

The box plot in Figure 16 is based on the so-called “quartiles,” which divide the distribution into four equally filled intervals. The upper and lower edges on the boxes are located at the first and third quartiles of the data, respectively (Jacoby, 1998). The vertical lines above and below the boxes (so-called “whiskers”) extend to the upper and lower adjacent values, or values that correspond to the maximum and minimum values in the data set. It can be seen in Figure 16, for example, that the mean *F*1 computed from the recordings obtained with the lavalier microphone has the shortest whiskers, indicating a relatively small distance between the highest and lowest *F*1 frequency in the data set. The horizontal line inside each box is the median line. If the line is off-center, the plot indicates an asymmetrical density of data points. Figure 16, for instance, shows that the

<sup>7</sup> With this type of between-subject follow-up test, it might be argued that the ANOVA has to be significant at the .025 level.

distribution of  $F1$  values obtained with the head-set microphone is the most symmetrical, while the  $F1$  distribution obtained with the built-in microphone is spread over the broadest range of values. At the same time, none of the methods show any values beyond the adjacent values, which indicates that no “unusual data” or outliers have been found.

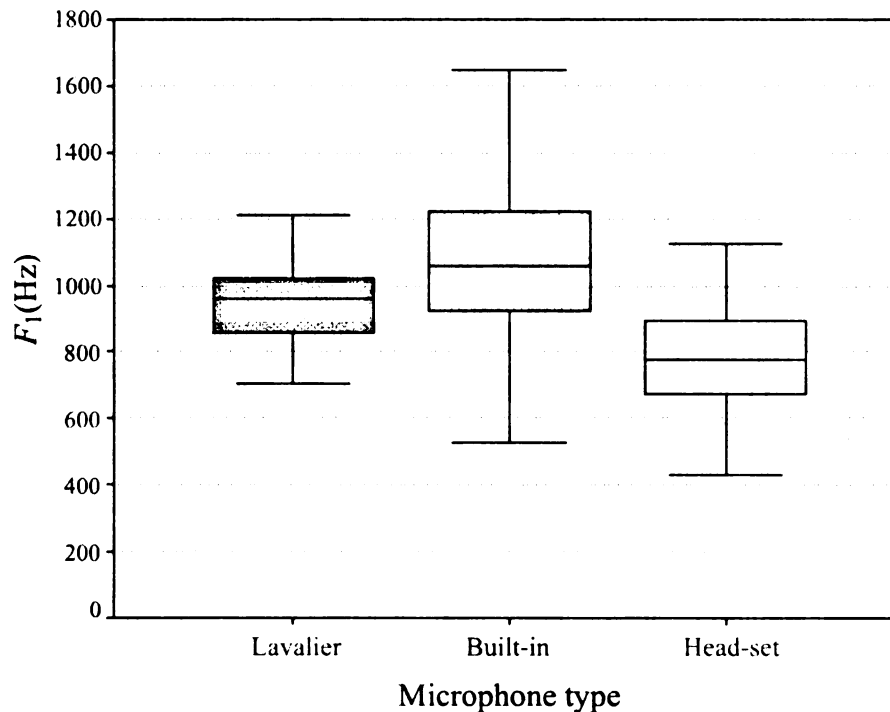


Figure 16. Box plot of overall between-subject comparisons of the data acquisition test

### 2.3.6.3 Summary of the signal acquisition experiment

As noticed in 2.3.6.2, the choice of data acquisition technology and methodology is an important consideration. Depending on the signal acquisition method used, significantly different results may be obtained across the  $F1$  domain. Researchers should be encouraged to follow best practices in the area of speech recording, processing, and analysis, and should use modern technology to their advantage (Plichta, 2002). There should also be no doubt that signals acquired with the head-set microphone and a 24-bit

digital recorder are the most accurate, and, therefore, the most reliable. Such signals contain the greatest amount of spectral detail and the least amount of unwanted noise.

The intent of this dissertation is not to advocate the abandonment of traditional methods, nor is it to claim that all traditional sociophonetic research is unreliable. On the contrary, there is an extremely rich sociophonetic tradition of acoustic analysis that has furthered our understanding of language variation and change probably more than any other branch of modern sociolinguistics. Still, NCCS-influenced vowels, due to their complex spectral features, are difficult to analyze, and caution must be exercised in this type of analysis.

To illustrate this point, Figure 17 shows three LPC spectra of the vowel /a/ in the word “lot,” obtained from the same corpus. LPC was applied at the same point in time and with the same parameters (analysis width of 0.025s, the sample rate of 16,000 Hz, LPC filter order of 13, and a mid-range pre-emphasis filter starting at 50 Hz). The vowel /a/, one of the major participants in NCCS, is problematic for an LPC algorithm because  $F1$  and  $F2$  are quite close to each other in frequency.

In poor quality, highly attenuated recordings (such as that in Figure 12), the distinction between the two peaks is blurred and LPC can return an incorrect reading. Figure 17 shows three LPC filter response graphs superimposed on top of one another. The box below the LPC graph contains the windowed (0.025s) portion of the waveform. The dashed line represents the sample obtained with a built-in microphone, and, as can be seen, it captures only one peak in the vicinity of  $F1$ . The same is true of the sample recorded with the lavalier microphone (dotted line). It is only the recording obtained with

the head-set microphone that allows the LPC filter to return two formant values and a very realistic-looking spectral envelope. There are many examples of this kind, and the data acquisition test carried out earlier in this chapter supports this point as well.

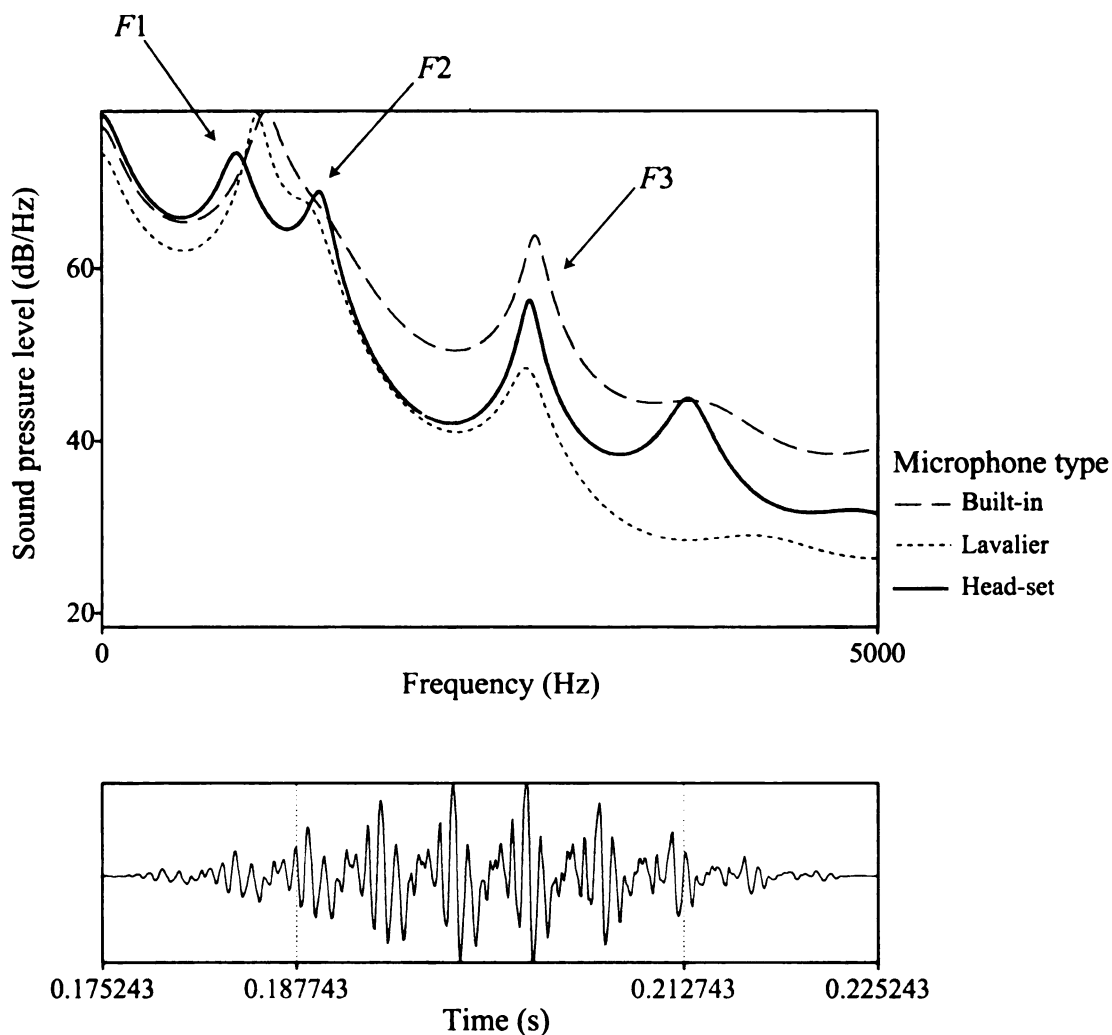


Figure 17. LPC spectra of the /a/ vowel in "job" by a female talker with an NCCS-influenced vowel system acquired with by different methods

### 2.3.7 LPC analysis of nasalized vowels: Is /æ/ really raising?

Traditional acoustic analysis methodologies may have lead researchers to believe that the vowel /æ/ in NCCS-influenced speech is significantly raised, while, it can be argued that

the reported, extremely low  $F1$  values, normally associated with increased vowel height, are a result of a combination of instrumentation limitations and the application of LPC (Linear Predictive Coding) acoustical analysis to nasalized spectra.

Numerous studies have reported  $F1$  values of /æ/ to be as low as 350-400 Hz (especially in female talkers), particularly when followed by a nasal consonant (e.g., Labov et al, (1972) or Labov et al, (1997)). Figure 18 shows two vowel systems obtained from female Detroiters (Labov et al., 1997). Each of these systems shows significantly raised /æ/, whose  $F1$  is considerably lower in frequency than that of /ε/. Most typically, such results have been obtained from tape-recorded (analog cassette or MiniDisc with an omnidirectional condenser microphone) or telephone-recorded wordlist and conversational data. Linear Predictive Coding (LPC) has most often been employed to extract formant values (e.g., Labov (2001)).

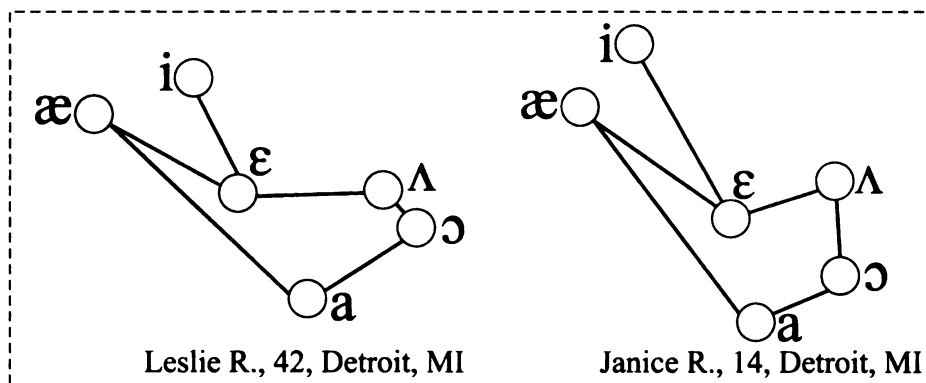


Figure 18. Sample vowel systems of Detroit, MI females (from Labov et al. (1997) with permission)

LPC models the vocal tract as an all-pole filter and does not account for spectral zeroes (anti-formants), which are present in nasalized speech sounds. In the presence of nasal



formants, even a small incorrect adjustment of the LPC filter order (a coefficient based on the number of expected formants in a specific digital audio file) might easily return incorrect formants values (e.g., a low  $F1$  of /æ/).

While it is not always possible to review and re-analyze audio data used in previous research, one might speculate that the application of LPC to nasalized vowels (e.g., /æ/ in the word “man”), particularly with poorly recorded data, might have misrepresented the picture of the /æ/ vowel’s movement within the vowel space, at least to some extent.<sup>8</sup>

Further in this chapter, it will become clear why this is might be the case.

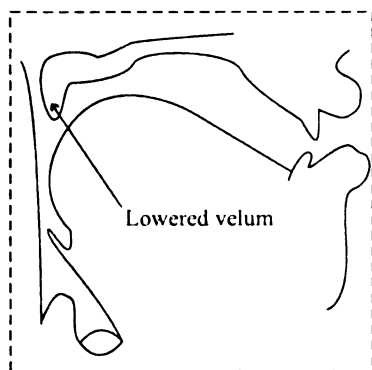
## **2.4 VOWEL NASALIZATION**

### **2.4.1 The velopharyngeal port**

Vowel nasalization can be generally of two kinds: (1) contrastive (phonemic) and (2) non-contrastive (allophonic or coarticulatory). In each case nasalization is caused by the opening of the velopharyngeal port as shown in Figure 19.

---

<sup>8</sup> Note that the vowel /ɛ/ is susceptible to the same problems, except that it has been reported to be lowering, not raising.



**Figure 19. Schematic view of lowering (opening) of the velum during the production of nasalized vowels**

The total area of the opening of the velopharyngeal port varies greatly among languages and dialects, thus causing different degrees of vowel nasalization. The opening of the velopharyngeal port creates an air passage between the oral and nasal cavities. This cavity coupling changes the spectrum of speech sounds produced in this way. Figure 20 shows a schematic, and simplified, view of the physiology involved during the production of nasalized vowels. In normal speech, the velum does not completely close the pharyngeal passage to the nose, and the degree of opening is subject to coarticulatory variability.

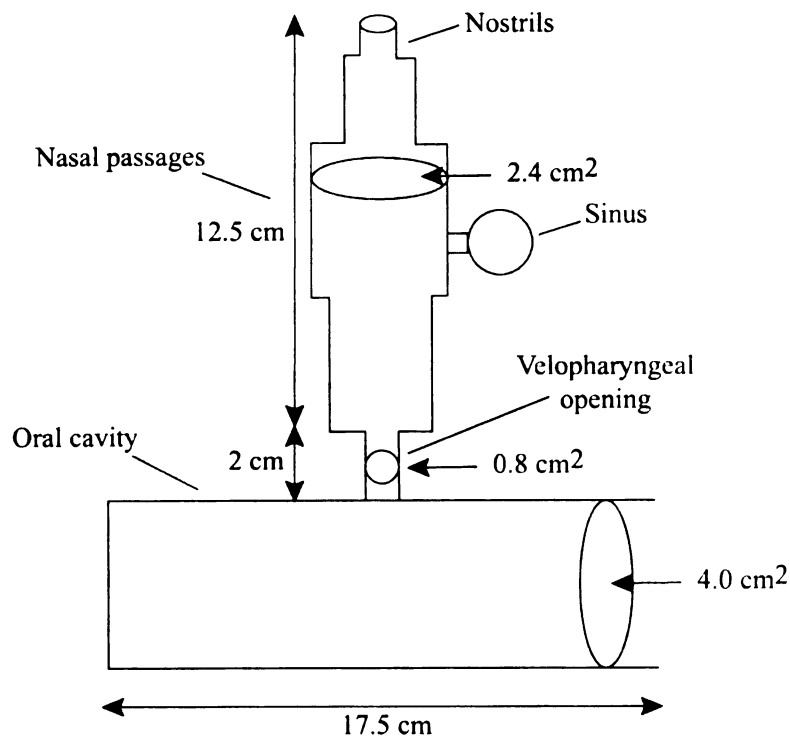


Figure 20. Schematic view of the oral and nasal passages (of a male talker) involved during the production of nasalized vowels (based on Chen (1997))

### 2.4.2 Oral and nasal formants

Just as the oral cavity produces its own resonant frequencies, so does the nasal cavity. For instance, the spectrum of the nasalized vowel / $\tilde{a}$ / produced with oral and nasal resonances reveals the presence of two extra peaks (nasal formants): one nasal formant is typically between the first two oral formants ( $F2_n$ ) and one at much lower frequencies, below the first formant ( $F1_n$ ).<sup>9</sup> At the same time, the bandwidth of  $F1$  is significantly broadened (to around 200-300 Hz) and there is evidence of two additional zeroes in the spectrum. Nasal formants and zeroes (or anti-formants) typically occur together as pole-zero pairs (e.g., Chen (1997)). The relative frequency and amplitude of the nasal formants

---

<sup>9</sup> Note that  $F2_n$  is sometimes referred to as " $F_n$ " in the literature.

depend on the area of the opening of the velopharyngeal port. Figure 21 contains a theoretical model of this phenomenon. Note that the first formant of the nasalized vowel ( $F1_n$ ) is lower in frequency than the first oral formant ( $F1$ ) and that the extent of this increases with the increase in velopharyngeal opening.  $F_n$  is the higher frequency nasal formant (in this dissertation, referred to as  $F2_n$ ) and  $f_z$  is a nasal zero (anti-formant). Again, note how all of the above parameters change as the opening of the velopharyngeal port increases (also see 2.4.3).

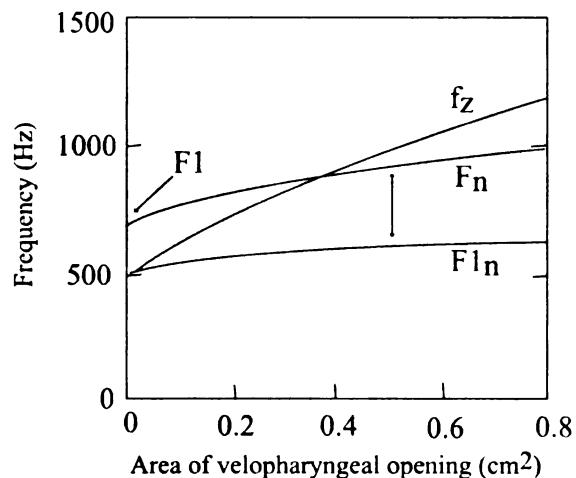


Figure 21 A theoretical model of the relationship between the area of the opening of the velopharyngeal port and the frequency and amplitude of oral and nasal formant for the vowel /æ/ (from Stevens (1998))

### 2.4.3 Spectral characteristics of synthetic nasalized vowels

In order to illustrate nasalized spectra in more depth, a series of four variants of the word “sack” has been synthesized with increasing degrees of opening of the velopharyngeal port according to the model proposed by Stevens (Figure 21). In Figure 22, the velopharyngeal port opening is 0 mm<sup>2</sup>, (left), and 7 mm<sup>2</sup>, (right). In Figure 23 it is 14 mm<sup>2</sup> (left) and 21 mm<sup>2</sup> (right). As the opening of the velopharyngeal port increases, so

does the amplitude of the nasal formant  $F1_n$ . The spectral envelope acquires a different shape with a great deal of energy close to the first harmonic and a significantly broadened bandwidth and reduced amplitude of the oral  $F1$ .

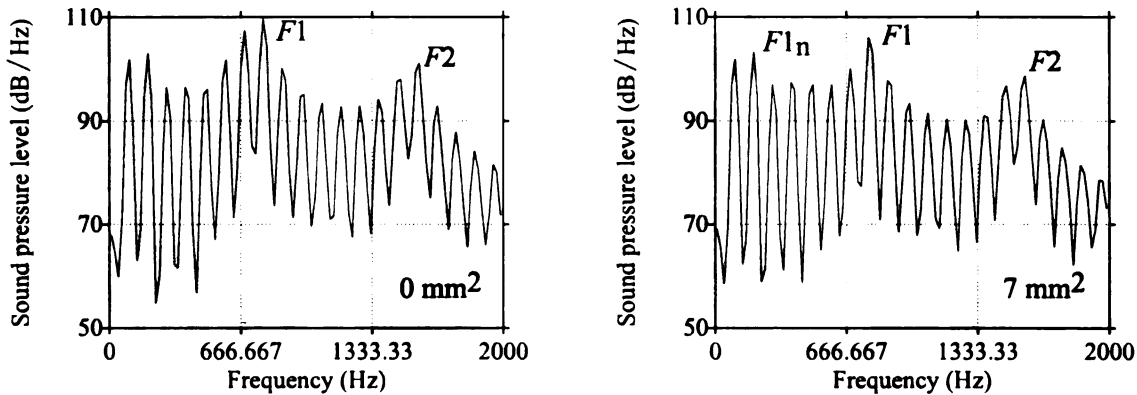


Figure 22. Spectrum of the /æ/ vowel synthesized with the velopharyngeal port opening of  $0 \text{ mm}^2$ , (left), and  $7 \text{ mm}^2$ , (right)

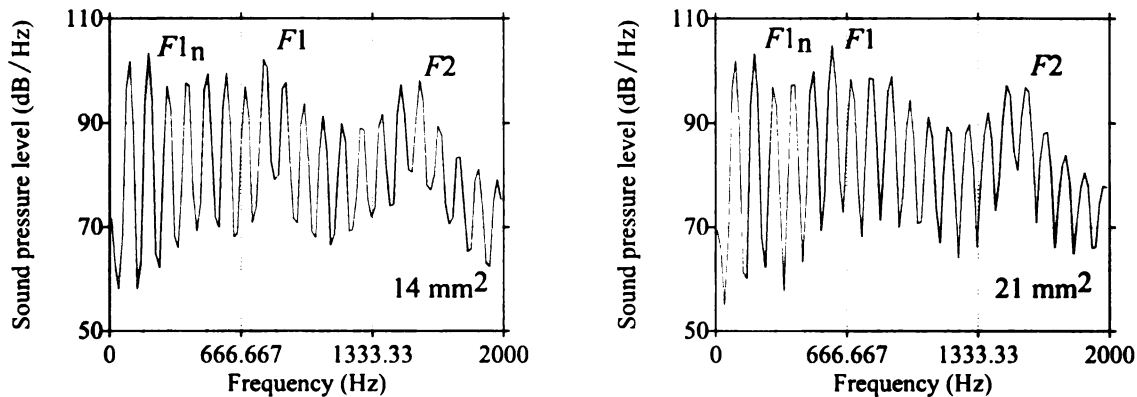


Figure 23 Spectrum of the /æ/ vowel synthesized with the velopharyngeal port opening of  $14 \text{ mm}^2$  (left) and  $21 \text{ mm}^2$  (right)

#### 2.4.4 Spectral characteristics of the oral ~ nasal contrast in Polish

As mentioned in section 2.4.1, vowel nasalization is used contrastively in some languages. Polish is one such language, and it has two nasal vowels: / $\tilde{e}$ / and / $\tilde{o}$ /. Figure 24 shows a spectrogram of the minimal pair /kɾeta/ ~ /kɾ $\tilde{e}$ ta/. The spectrogram of /kɾeta/

shows well-defined, high-amplitude formants with narrow bandwidths (the dark bands in the illustration), while the spectrogram of /kr̥ɛ̃ta/ shows a great deal of energy diffusion around  $F1$ , as well as the presence of the nasal formant  $F1_n$ .

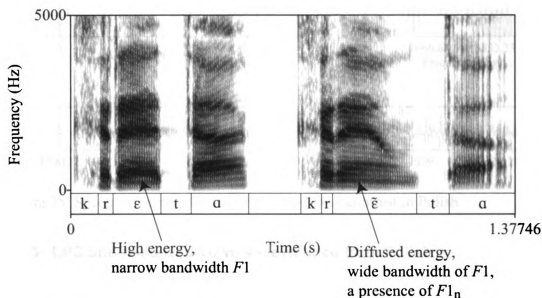


Figure 24. Spectrogram of a Polish minimal pair /kreta/~/kr̥ɛ̃ta/

Figure 25 shows a spectral “slice” taken at the mid-point of each vowel nucleus. There is no evidence of velopharyngeal opening in the non-nasal example.  $F1$  has a well-defined peak, and the spectral envelope shows an approximately -6 dB/octave sloping pattern characteristic of this mid-front vowel. The  $F1$  peak is strong and has a narrow bandwidth. The spectrum of the nasal vowel /ɛ̃/, on the other hand, looks quite different. There is a high energy concentration around the second harmonic – evidence of the low-frequency nasal peak. Also, the  $F1$  is lower in amplitude and its bandwidth is significantly wider than for the non-nasal vowel /ɛ/.

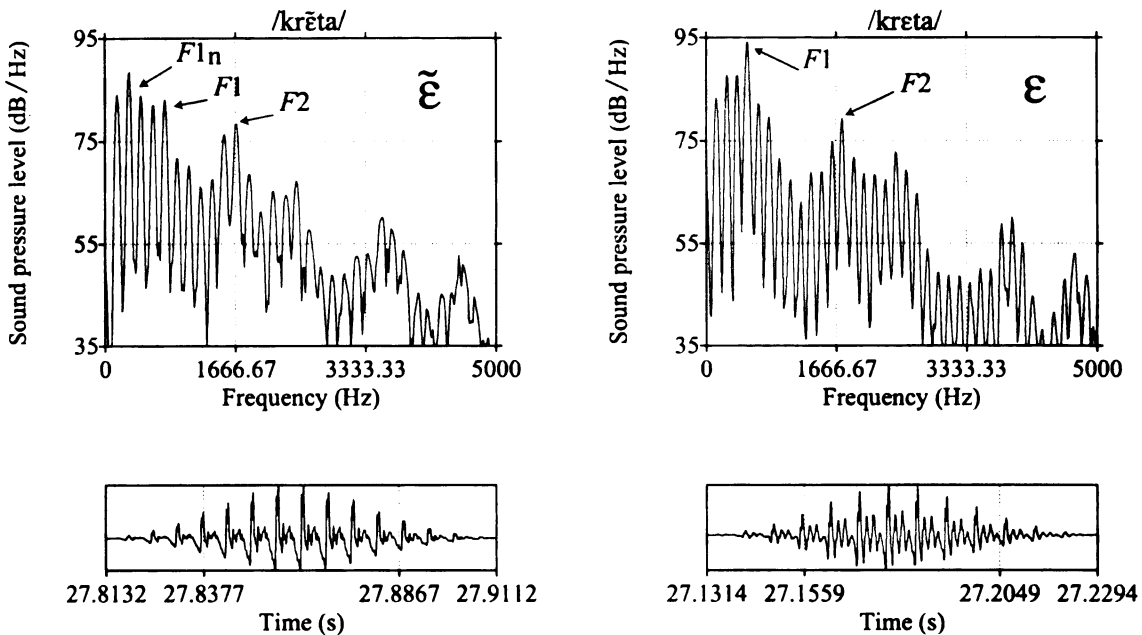


Figure 25. Spectral characteristics of the oral ~ nasal contrast in Polish

#### 2.4.5 LPC and nasalized vowels – evidence from Polish

Linear Predictive Coding (LPC) can be thought of as an all-pole digital filter with a response similar to that of human speech. LPC has been used with a great deal success in various computational areas – from speech recognition to digital watermarking (e.g., Gurijala, Deller, & Seadle (2002)). LPC has also been used in sociophonetics and has constituted the predominant technique for extracting formant values from field-recorded speech (e.g., Labov (2001)).

However, LPC does not work well with nasalized vowels. LPC algorithms cannot account for the additional nasal formants and zeroes. This, coupled with poorly acquired, highly attenuated, noisy recordings can result in the return of incorrect formant values from LPC analysis. Figure 26 shows LPC and FFT spectra of the Polish minimal pair /p̥ɛta/~pɛta/. In the nasal case (left panel), close to the LPC-returned peak labeled “F1”,

a strong nasal peak around the second harmonic can be seen. This peak is followed by three harmonics that are very close to one another in amplitude. At the same time, the second nasal peak can be found approximately half-way between the LPC peaks labeled “ $F1$ ” and “ $F2$ ”, and an additional anti-formant is predicted to be found in the area just above 1,000 Hz (see Figure 21 for comparison). The LPC algorithm is not designed to handle vowels with such complex spectra and it returns  $F1$ ,  $F2$ , and  $F3$  values that seem to be averaged over the oral and nasal peaks and zeroes. By contrast, the spectrum of the oral vowel / $\epsilon$ / (right panel) is not as problematic for the LPC algorithm, which returns realistic approximations of formant values.

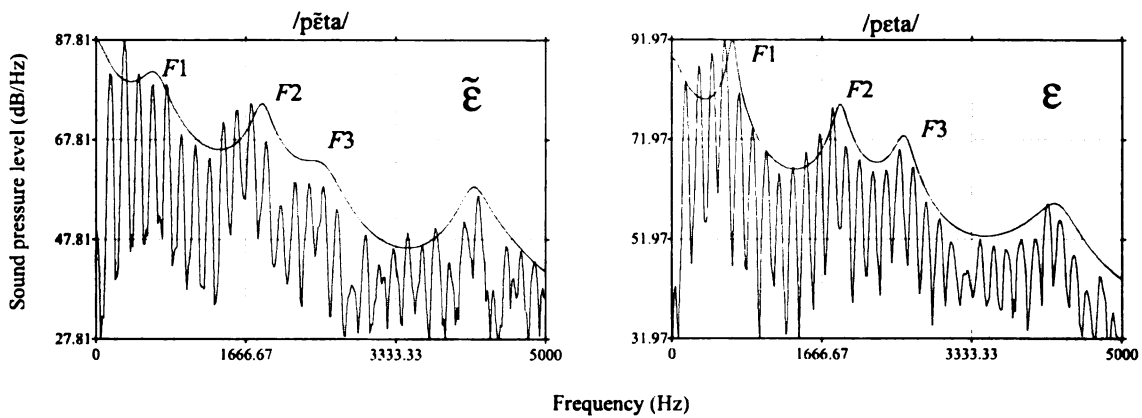


Figure 26. Comparison of LPC and FFT spectra of / $\epsilon$ / and / $\tilde{\epsilon}$ /



## **2.5 VOWEL NASALIZATION IN MICHIGAN**

### **2.5.1 Nasal formants appear in the spectrum of Lower Michigan vowels**

Careful acoustic analysis of certain NCCS-influenced vowels, particularly /æ/, in non-nasal environments<sup>10</sup> reveals evidence of two nasal formants. In the case of /æ/, for example, one nasal formant,  $F2_n$ , is usually found between the first two oral formants ( $F1$  and  $F2$ ) and one,  $F1_n$ , is found below the first formant. At the same time, the bandwidth of  $F1$  is significantly increased, relative to vowels showing no signs of nasalization. This makes the vowel /æ/, for example, quite similar to both the theoretical model of a nasalized vowel, as well as the real-life Polish example cited in 2.4.5. Figure 27 shows two spectra of the vowel /æ/ in “back” - one without any evidence of velopharyngeal port opening (left), and one showing evidence of at least the first nasal formant,  $F1_n$  (right). The non-nasal /æ/ was obtained from a young African-American female speaker from Saint Paul, MN (a speech community not influenced by NCCS), while the nasalized sample was obtained from a young female talker from the Detroit area in Michigan (an area influenced by NCCS). The waterfall spectrograms in Figure 28 show the entire word “back” and demonstrate the presence of nasal energy in more detail than the spectral slices in Figure 27. Note high energy concentration and diffusion along the bottom of the spectrogram of /bæ̃k/ in Figure 28 (right), similar to that found in Figure 24.

---

<sup>10</sup> Non-nasal environments are defined here as the presence of oral consonants or vowels both preceding and following the vowel. For example, “hat.”

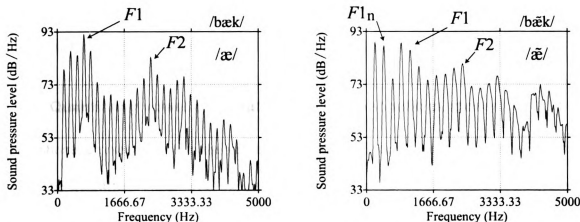


Figure 27 Examples of non-nasalized vowel spectra (left) and nasalized spectra (right.) of the vowel /æ/ in "back"

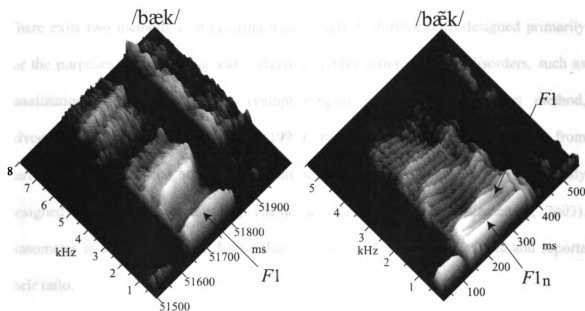


Figure 28. Waterfall plots of the word "back" of the same sample as in Figure 24

## 2.5.2 Why a sociophonetic aerodynamic analysis of vowel nasalization?

In light of the evidence of potential data acquisition problems and their impact on derived formant values, particularly  $F1$ , and because evidence of nasalized vowels has been found during routine sociophonetic analysis of NCCS-influenced speech, it was thought important to pursue an alternative diagnostic methodology to see whether nasalization is

regularly associated with NCCS vowels and, if so, how it patterns with sociolinguistic factors such as sex, and region.

### **2.5.3 Quantifying vowel nasalization for sociophonetic purposes**

Sociophonetic research within the variationist framework demands that variable phonetic details be quantified either with objectively measurable values or researcher-subjective indices (Chambers, 1995). Having established the presence of nasal cavity coupling in non-nasal environments in certain Detroit-area individuals, one must now devise an appropriate quantificational methodology to capture the social distribution of this feature. There exist two methods of measuring nasalization; both have been designed primarily for the purposes of diagnosing and correcting certain communication disorders, such as nasalization due to cleft palate (velopharyngeal inadequacy). The first method, advocated, among others, by Chen (1995), involves quantifying nasalization from narrow-band spectra; the other involves an aerodynamic measurement with a specially designed device, such as the Kay Elemetrics Nasometer II (KayElemetrics, 2003). Nasometer is a head mounted device that registers oral and nasal airflows and reports their ratio.

#### **2.5.3.1 Quantifying nasalization from narrow-band spectra**

Quantifying nasalization from narrow-band spectra has been proposed by several researchers but Chen (1995, 1997) presented the most compelling and comprehensive case. The method proposed by Chen is based on calculating amplitude ratios between oral and nasal peaks. In this framework, the amplitude of the nasal formant, which is most typically found between the first two oral formants, is called “P1” while the amplitude of

the nasal formant, typically found below the first formant, is called “P0.” Based on the theory of speech production (see, for instance, 2.4), Chen assumes that the amplitude of the oral  $F1$  (called “A1”) will be substantially reduced due to nasal cavity coupling. For Chen, this reduction in  $A1$  is essential to quantifying nasalization. Chen proposes calculating two variables (Figure 29):

1.  $A1-P1$  – the difference in amplitude (in dB) between the first oral formant (A1) and the second nasal formant (P1), and
2.  $A1-P0$  – the difference in amplitude between the first oral formant (A1) and the first nasal formant (P0)

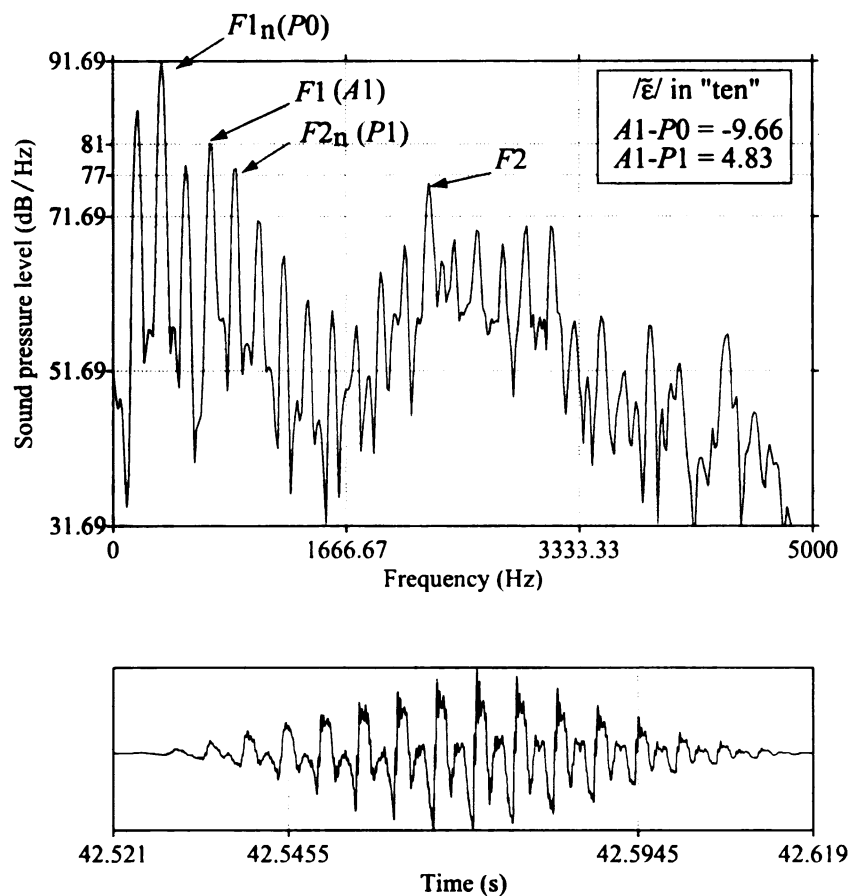


Figure 29. Oral and nasal prominences involved in Chen's method

Chen makes a strong case in favor of this approach and, from the theoretical standpoint, this approach shows great promise. However, problems are likely to occur for the following reasons:

1. It is not always obvious where the specific peaks are to be found. Vowels that have only a small amount of nasalization or none at all will be particularly problematic for this method.
2. This method is relative to the absolute formant frequencies of the vowel under investigation and it requires an extra vowel-type adjustment procedure that is necessary for finding the correct *P1* and *P0* peaks. Moreover, formant transitions due to the phonetic environment preceding and following the vowel would require adjustment as well.
3. This method only works when data are acquired by methods similar to those described by Chen – with a high-quality omni-directional microphone (presumably to avoid proximity effect) close to the talker's lips and in a sound-proof room. One might speculate that the output of this type of analysis would be strongly affected by the type of recording equipment used and the circumstances in which the recording was made. For example, most studio or stage directional microphones have purposefully emphasized low frequencies and have a high degree of the proximity effect, which is desired by stage performers and newscasters, but which can compromise the reliability of the Chen method (see 2.3.5 for a discussion on role of the microphone).
4. This method is not optimal for variationist work, as one has to assume a priori, where the nasal peaks are to be found (as a function of the velopharyngeal port opening, for example).

It is largely because of these reasons that this method was not employed in the present study as the main diagnostic tool. Instead, it was used as a secondary method.

### **2.5.3.2 Designing peak-seeking software for Chen's method**

Chen recommended looking for the peaks in the vicinity of the harmonics closest to the peaks predicted from theory. Chen does elaborate on the theoretical model herself, but a more detailed account can be found in Stevens (1998). The software would have to be

able to find the right peaks, verify them, get their frequency and amplitude values, and compute all required variables (Figure 30). The most important, and most difficult, part of the design process was to make sure that there was the least amount discrepancy between the peaks picked by software and those picked manually.

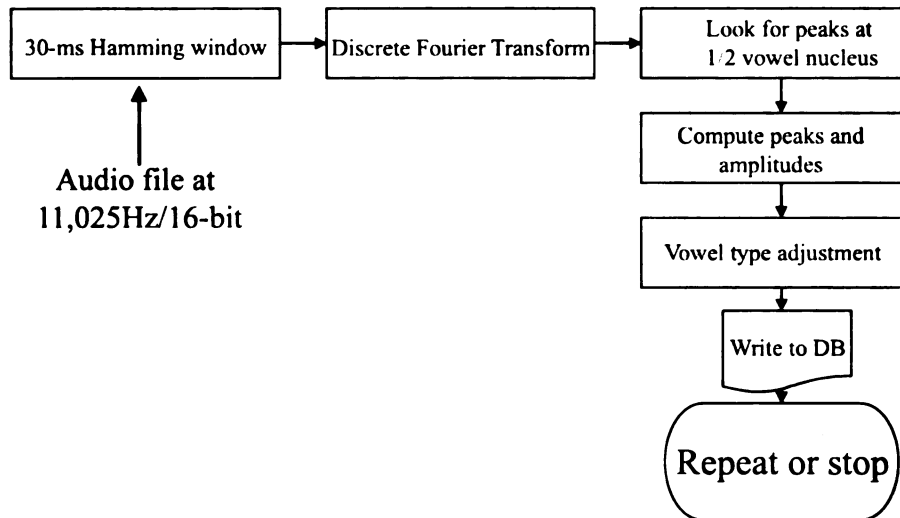


Figure 30. Flowchart of *Akustyk*'s algorithm designed to automate the peak seeking process

### 2.5.3.3 Successes and failures of the Chen method

As applied here (see 2.8 for details), Chen's method proved moderately successful. The criticism about finding low frequency peaks turned out to be well-founded. In tests on vowels recorded with different microphones, and on vowels recorded with low frequency roll-off filters (very common on many preamplifiers and microphones), *A1-P0* was not reliable, as the variable low-end part of the spectrum considerably biased the results (see Figure 14). In such situations, one is faced with a difficult question of whether the diagnostic method is flawed or whether it is the investigator's responsibility to acquire pristinely flat recordings. In this particular case, the method's special sensitivity to low

frequency fluctuations made it very difficult to acquire reliable signals. In practice, only the head-worn Sennheiser HMD 25-1 and the Earthworks M30BX measurement microphone captured the speech signal without significant low-end bias. Shure SM58, Audio-technica A831b, and Audio-technica AT822 were also tested but proximity effect compromised signal reliability.

On the other hand, the  $A1-P1$  variable worked out relatively well, as the value of  $P1$  was much less likely to be affected by low frequency bias, and the difference between manual and machine-obtained values was within a 10% tolerance. Perhaps the biggest problem with  $A1-P1$  was that for some female talkers, the difference in frequency between  $A1$  and  $P1$  peaks was very small, particularly for the vowel /a/. Thus, for instance, the  $F1$  of /a/ will come very close in frequency and amplitude (e.g., neighboring harmonics) to the predicted nasal peak at 950 Hz, and the difference in amplitude between these two peaks will be very small. As a result, a high level of nasalization (i.e., small  $A1-P1$ ) will be reported while the sample might not be nasalized at all (see a discussion on vowel type adjustments in 2.8).

#### **2.5.3.4 Quantifying nasalization from aerodynamic data**

Quantifying nasalization with a Nasometer-type device is free from the problems of Chen's method. The method is based on a very simple, yet robust design. Two transducers are placed in two chambers separated from each other with a sound attenuating barrier. Each transducer signal is fed into a separate channel of a stereo (unbalanced) mini-plug connector. The microphones require a constant power supply (also known as "phantom power"), so there must be a power source available – typically

a AA battery. The device connects to a stereo, line-level interface of a common 16-bit sound card such as SoundBlaster 16.

Software captures the signal coming through the stereo channel, splitting it into the nasal and the oral signals – left and right. It then performs a series of computations to obtain several common acoustic parameters, such as  $F_0$ ,  $F_1$ , and amplitude. The oral-to-nasal energy ratio is then calculated and reported in real time. Partly because of the simplicity of the design, and partly because of the availability of very sensitive transducers and good software to process the data, this method has the promise of becoming a good tool for measuring vowel nasalization for variationist purposes. The only notable problem is that the standard nasometers, such as the Kay Elemetrics Nasometer II, are quite bulky, heavy, and obtrusive.

While Nasometer II may work quite well in a clinical setting, it is generally difficult to use in the field. There is one type of nasometer, though, that has all the benefits of a stationary one, but is small, portable, and unobtrusive. It is the Rothenberg mask (Rothenberg, 1995). The Rothenberg mask (Figure 31) is made of a synthetic material and is shaped to fit the human face, though it probably should not be used with persons with a lot of facial hair.





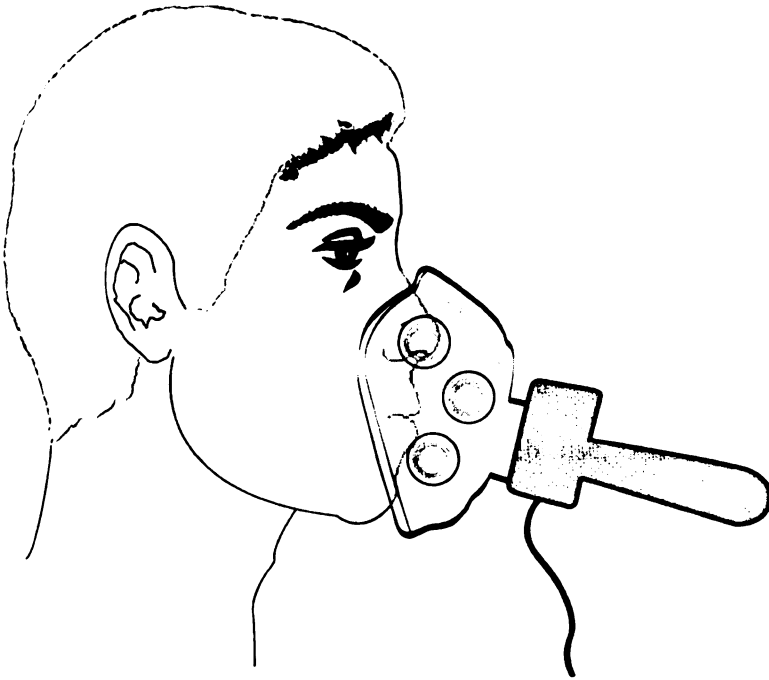


Figure 31. The Rothenberg mask

The oral and nasal chambers are divided with a thick plastic barrier in order to minimize sound leakage between them. There are also small, specially designed circular vents placed in the mask to facilitate breathing and prevent muffling while wearing the mask. The Rothenberg mask works almost exactly like other nasometers, but it has the added benefit of calculating %Nasalance (also known as “*F0* nasalance”) from the amplitudes of only the fundamental frequency components of the nasal and oral acoustic energy.

***F0* nasalance** is less sensitive to specific vowel formants and voice pitch than is ***F1* nasalance**, and is less sensitive to articulatory movements than are methods comparing **low-pass filtered** airflows. The parameter that will be used in the subsequent analysis is **often referred** to as %Nasalance, or %N. These two terms will be used interchangeably in **this dissertation**. %N is calculated from the energy measured at the nares ( $A_n$ ) and the **energy measured** at the mouth: ( $A_o$ ).

$$\%N = \frac{A_n}{A_o + A_n} 100 \quad (2)$$

### **2.5.3.5 Weaknesses of the airflow methodology**

An obvious disadvantage is the need for the subject to wear a mask. However, thanks to the circular wire vents, and a very lightweight, small form factor design, the mask did not cause any subjects in the present study significant problems. One must remember, however, to disinfect the mask after each use. Another obvious limitation is the possibility of sound leakage between the oral and nasal chambers. This problem is generally not very serious. However, it does intensify when there is complete velopharyngeal closure and a very strong oral constriction. This is why the non-nasal high vowels /i/ and /u/ may appear to have higher %N than mid or low vowels. This is, of course, contrary to the physiology of speech production. This problem occurs because the high oral energy leaks into the nasal chamber and thus registers as high nasalance (also see 2.6.4 for a detailed discussion).

## **2.6 STUDY DESIGN AND METHODS**

### **2.6.1 The goals**

*The present study* investigated the degree of vowel nasalization across three previously *studied dialectal* groups in Michigan: European American Detroiters (“LM”), European *American rural* Mid-Michiganders (“MM”), and European American Upper Peninsula *talkers* (“UP”). The dialects of Detroiters and Mid-Michiganders have been well *documented* (Labov et al., (1972) and Ito, (2000) respectively), as have other groups *participating* in NCSS (Evans, 2001).

For example, it has been established that young, middle-class and upper-middle-class, socially mobile individuals are most likely to exhibit advanced stages of NCCS. In addition, women are often ahead of men in adopting these types of chain-like, vocalic changes (Labov, 2001). The goal of this study was, therefore, to investigate the populations identified in previous work as likely “vowel shifters” and to see whether nasalization, or %N to be exact, correlates with their shifted phonology. It was not the goal of this study to demonstrate a detailed social stratification of Michigan residents with regard to vowel nasalization.

### **2.6.2 The subjects**

The studies by Ito (2000) and Evans (2001) demonstrated that the adoption of NCCS in Michigan is gradient with regard to region and sex. In order to study the nasalization effect as a potential correlate of NCCS, it was essential to select subjects from demographic groups that are likely to have vowel systems influenced by it. The sample for this study was, therefore, drawn from among three distinct geographical areas – South-Eastern Michigan (n=12), Mid-Michigan (n=8), and the Upper Peninsula (n=10). The subjects were young (19-35) and of a middle-class or upper middle-class background; all with relatively loose social networks in ways described by Milroy (1980). The subjects were screened for and free of the history of cleft palate, chronic sinus problems, chronic respiratory problems, smoking, facial hair, and common speech impediments.

### 2.6.3 Data collection

The subjects were asked to provide demographic information based on a sociolinguistic questionnaire (Appendix B). The answers were entered into an electronic database. Next, the subjects were recorded saying 50 monosyllabic target words, selected to offer a broad range of vocalic and consonantal contexts (Appendix A)<sup>11</sup>. The recordings were gathered with a Sennheiser HMD 25-1 microphone and a 24bit/48KHz digital hard disk recorder to obtain the highest quality digital representation of the acoustic signal, and to ascertain a flat frequency response.

Finally, the subjects were recorded saying the same 50 words by means of the Rothenberg mask (1995). The subjects were instructed on how to hold the mask in order to minimize sound (and airflow) leakage and attenuation. The words were uttered one at a time, with a 10 second pause in between to minimize potential coarticulatory effects across word boundaries. The signal was recorded at the sample rate of 22,050 Hz and at a 16-bit bit-depth. Since most nasal energy is concentrated in the lower bands of the spectrum, the effective frequency response of 11,025 Hz at 16-bit was sufficient to capture the signal with a high degree of accuracy. The samples were acquired via a stereo channel, with one channel for the oral, and one for the nasal data stream. Figure 32 shows spectra from the oral channel (left) and the nasal channel (right) of the vowel /a/ in the word “lot” by a female Detroiter.

---

<sup>11</sup> Target words with the high vowels /i/ and /u/ were not included in the corpus (see 2.6.4)

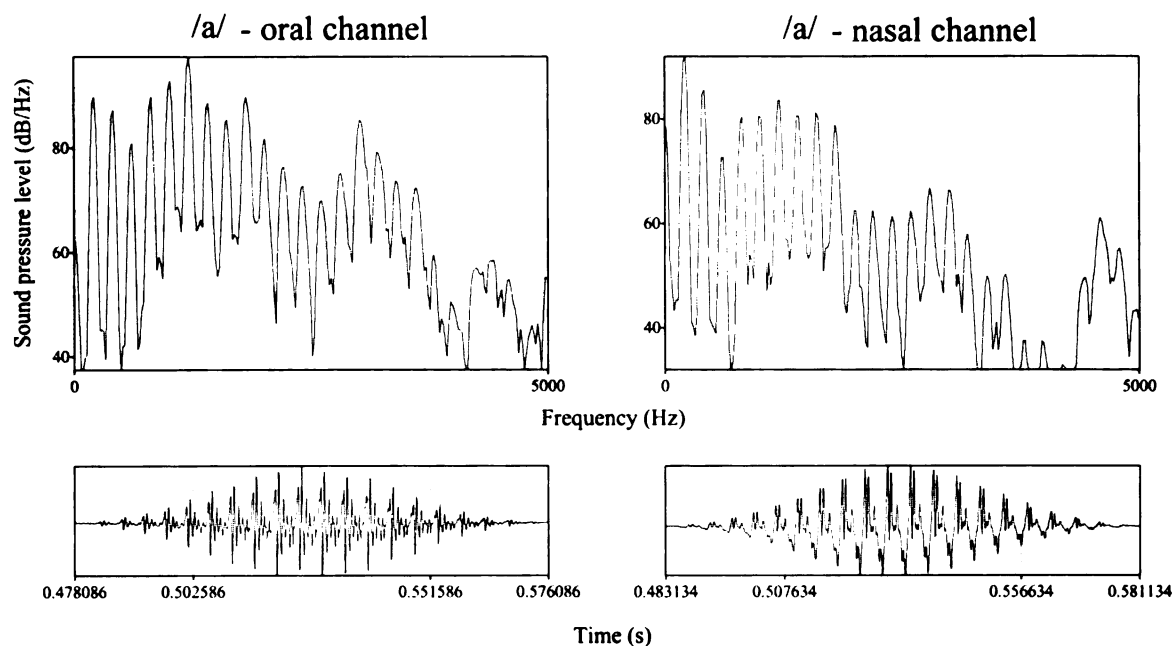


Figure 32. Spectra from the oral and nasal channel of /a/ in “lot”

#### 2.6.4 Why /i/ and /u/ were excluded from the study

There are two reasons why the high vowels /i/ and /u/ were excluded from the corpus:

1. Relatively low nasalization levels in the high vowels /i/ and /u/ are predicted on the grounds of experimental studies, such as Moll (1962) and Krakow and Huffman (1993) who demonstrated on x-ray and MRI images, respectively, that the velum is raised during the production of high vowels in general.
2. There is a reported energy leakage from the oral to the nasal chamber during the production of high vowels when there is both a complete closure of the velopharyngeal port, as well as a strong oral constriction anterior to the velum (Rothenberg, 1999). One would have been faced with a paradoxical situation – the vowels that are predicted to be the least nasalized would have registered a relatively high %N.

In addition, the vowels /i/ and /u/ have not been reported to be active in NCCS, unlike some other dialects of English (for instance, the Southern Vowel Shift) where /u/-fronting

is an active process (Fridland, 1998). Thus, the decision to leave them out of the analysis should be of little consequence to the overall findings.

## **2.6.5 Data processing**

### **2.6.5.1 Audio data**

The original 24-bit audio recordings to be used in spectral analysis were downsampled to 11,025 Hz/16-bit. They were saved in lossless PCM (Pulse Code Modulation) format. This made the data suitable for detailed acoustic analysis. A thorough acoustic analysis was performed by means of *Akustyk* (Plichta, 2004), written specifically to obtain a comprehensive spectral analysis of vowels. *Akustyk* offers advanced analysis tools, such as “on-the-fly” regression analysis of formant trajectories. It also keeps track of all acoustic and demographic data in a standard, SQL-compliant (Standard Query Language) format. This minimizes the risk of misplacing or mislabeling data before it reaches analysis software, such as Systat. *Akustyk* also allows statistical analyses of formants, such as paired or independent samples T-test, discriminant analysis, and Principle Component Analysis (see 2.9.1.2 for more information on formant extraction with *Akustyk*).

### **2.6.5.2 Nasalance data**

Data acquired with the Rothenberg mask did not require any additional signal processing. The data, already stored on a workstation hard drive, were analyzed by means of the OroNasal Mask System software (Glottal, 2002) and the parameter “% Nasalance” (%N), derived from the ratio of oral to nasal energy at  $F_0$  (see 2.5.3.4 for a definition of %N) was calculated and entered into a database.

### **2.6.5.3 Nasalization data from spectra**

As mentioned in 2.5.3.2, the Chen method (measuring nasalization from spectra) was automated with *Akustyk*. The software analyzed the entire corpus in batch mode. All analysis data were written to a database in real time.

### **2.6.5.4 Database queries**

After all the data had been collected, computed, and written to the database, a query was generated to merge the demographic data with the acoustic data and the nasalance data.

### **2.6.5.5 Useable data**

Due to the more difficult nature of acquiring data with the Rothenberg mask and writing it to a computer hard drive in real-time, there were slightly fewer useable %N tokens (1382) than spectral tokens (1418).

## **2.7 STATISTICAL ANALYSIS OF %N**

The null hypothesis associated with this study was that nasalization (%N) would be idiosyncratic and randomly distributed with respect to NCCS. However, if the qualitative spectrographic analysis in 2.5.1 was correct, then there would be a substantial relationship between NCCS and coarticulatory vowel nasalization. To test these hypotheses, the following specific research questions were formulated:

1. Will Lower Michigan respondents have a different level of nasalance than Mid-Michiganders and UP respondents?
2. Will women have a different level of nasalance than men?



3. Will the vowels involved in NCCS (/a/, /æ/, or /ε/) be specifically related to nasalization?

### 2.7.1 %N - two-way analysis of variance

The present study involved one continuous dependent variable, %N, and two factors: talker sex and region (on three levels: LM, MM, and UP). The first tests conducted were the omnibus, or overall, tests of the main and interaction effects of region and sex.

The test of between-subjects effects showed that region,  $F(2,1377)=164.536$ ,  $p<.001$ , sex,  $F(1,1377)=224.010$ ,  $p<.001$ , and their interaction,  $F(2,1377)=11.388$ ,  $p<.001$ , were all significant. Figure 33 shows overall means of % Nasalance across region and sex. Because the interaction effect was significant, simple main effects tests and interaction comparisons were performed. Their results are described below.

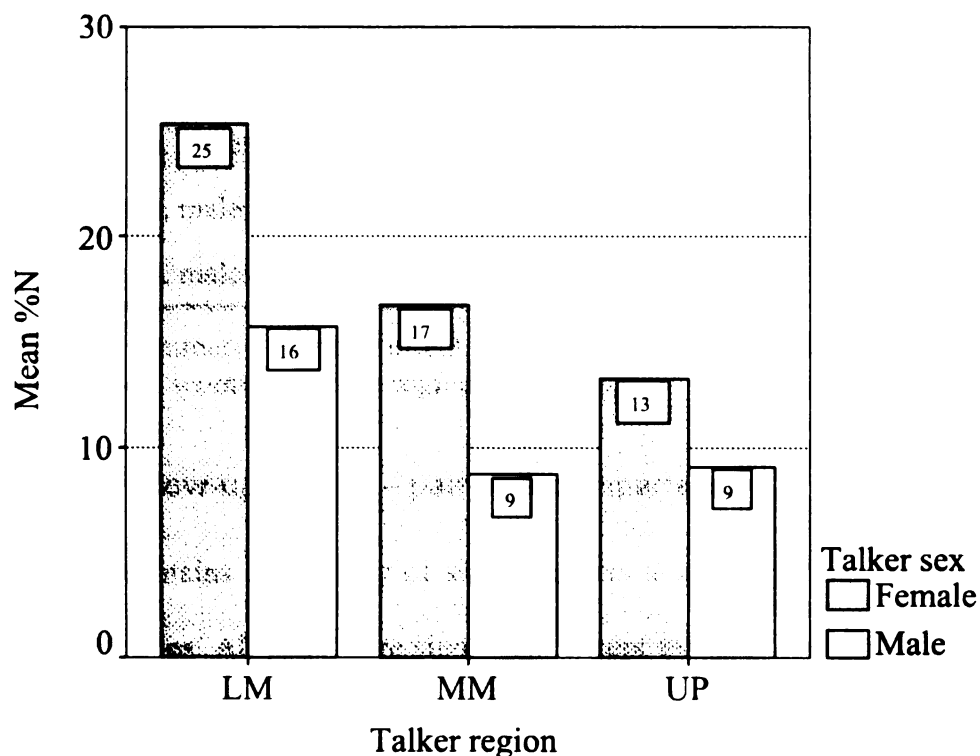


Figure 33. % Nasalance means by respondent region and sex

### 2.7.2 Follow-up test #1 – simple main effects tests

Because of the significant interaction effect found in 2.7.1, a follow-up test was conducted to answer two questions that have a great deal of importance in sociophonetic research:

1. Are the %N means across respondent regions significantly different for men and women separately?
2. Are the %N means within respondent regions significantly different for men and women for each region separately?

Table 2 contains a summary of answers to questions 1 and 2 above based on a family of main effects tests.

Simple effect tests for men and women across the three regions	
Females across regions	$F(2,1377)=155.616, p<.001$
Males across regions	$F(2,1377)=45.707, p<.001$
Females vs. males within LM	$F(1,1377)=179.579, p<.001$
Females vs. males within MM	$F(1,1377)=79.514, p<.001$
Females vs. males within UP	$F(1,1377)=21.792, p<.001$

Table 2. Summary of simple effect tests for men and women across the three regions and for men and women for each region separately

### 2.7.3 Follow-up test #2 – pairwise comparisons

Table 3 contains a summary of simple effects tests comparing different regions within sex categories. All pairs were significant except for male respondents between MM and UP.



Simple effect tests for men and women across each region separately	
Females between LM and MM	$F(1,1377)=117.971, p<.001$
Females between LM and UP	$F(1,1377)=292.486, p<.001$
Females between MM and UP	$F(1,1377)=17.22, p<.001$
Males between LM and MM	$F(1,1377)=71.407, p<.001$
Males between LM and UP	$F(1,1377)=54.369, p<.001$
Males between MM and UP	$F(1,1377)=.155, p=.694$ (ns)

Table 3. Summary of simple effect tests for men and women across each region separately

#### 2.7.4 Follow-up test #3 - interaction comparisons

In addition to simple main effect analyses, an interaction comparison was conducted. This test consisted of three tetrad comparisons to see whether the differences in %N means across the regions were the same or different for male and female respondents. Table 4 summarizes the results.

Differences in %N means across the regions	
LM vs. MM for females vs. LM vs. MM for males	$F(1,1377)=1.922, p=.166$ (ns)
LM vs. UP for females vs. LM vs. UP for males	$F(1,1377)=22.52, p<.001$
MM vs. UP for females vs. MM vs. UP for males	$F(1,1377)=9.1942, p<.025$

Table 4. Summary of three tetrad comparisons to evaluate whether the differences in %N means across the regions were the same or different for male and female respondents

#### 2.7.5 Summary of the statistical analysis of % Nasalance

Based on the statistical tests so far, as well as the nature of the data, the main findings can be summarized as follows (also see **Error! Reference source not found.**):

1. Women have significantly higher levels of nasalization than men across all regions.
2. Women have significantly higher levels of nasalization than men within each region.
3. Nasalization levels for women increase significantly in a gradient manner from the Upper Peninsula, through Mid-Michigan, to Lower Michigan.
4. Nasalization levels for men do not increase in the same manner as those of women. The Lower Michigan group is ahead of both Mid-Michigan and the Upper Peninsula, but Mid-Michigan men and Upper Peninsula men do not differ from each other.

### **2.7.6 Is nasalization global or local?**

One of the most important questions about nasalization in non-nasal contexts is whether it is a global or a local phenomenon. In other words, does it affect specific vowels or is it generalized over the entire vowel inventory? As mentioned in 2.6.4, the high vowels had been left out of the analysis, but for the remaining vowels, %N turned out to be statistically significant for the group as a whole,  $F(7,1127)=2.349$ ,  $p<.05$ . Subsequent pairwise comparison tests showed that the only significant difference was between the vowels /æ/ and /ε/ for male talkers.

The above findings indicate that vowel nasalization in non-nasal contexts is a phenomenon generalized over the entire vowel inventory. Figure 34 shows mean %N values for men and women for non-high vowels. Also, note very similar distributions patterns of %N among male (gray bars) and female talkers (white bars).

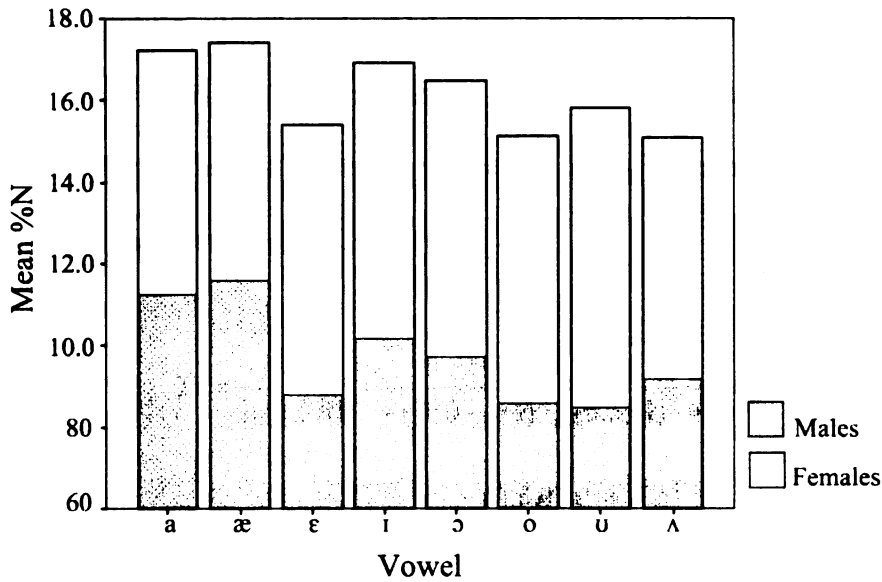


Figure 34. Male and female %N distribution patterns by vowel in non-nasal environments and for non-high vowels

## 2.8 STATISTICAL ANALYSIS OF A1-P1

As mentioned in 2.5.3.3, the Chen method proved to be relatively reliable for the computation of *A1-P1*. For reasons described in 2.5.3.3, this method was used as a secondary method, primarily to see whether measuring nasalization reliably from spectra in batch mode was possible. *A1-P1* was subject to the same type of statistical analysis as %N, both because of the nature of the data, and also to be able to compare the two methods more closely. The spectral method yielded very similar overall results to the aerodynamic method. Table 5 summarizes overall means and standard deviations. Note a rather high standard deviation for *A1-P1* obtained in the present study. Chen (1995) proposed a vowel type adjustment procedure to alleviate some of the vowel-specific problems listed in 2.5.3.1. The amplitude *P1* in *A1-P1* is to be replaced by (*P1-T1-T2*) in *dB* calculated according to the formulas below, where *T1* is the effect of the first formant on the extra peak due to nasal coupling, *T2* is the effect of the second formant



component,  $F1$  is the first oral formant frequency,  $F2$  is the second oral formant frequency,  $B1$  is the bandwidth of  $F1$ , and  $B2$  is the bandwidth of  $F2$ . The constant value of 950 Hz is derived from the theoretical model of the predicted frequency of second nasal formant (see Figure 29).

$$T1 = \frac{(0.5B1)^2 + F1^2}{[((0.5B1)^2 + (950 - F1)^2)((0.5B1)^2 + (F1 + 950)^2)]^{1/2}} \quad (3)$$

$$T2 = \frac{(0.5B2)^2 + F2^2}{[((0.5B2)^2 + (F2 - 950)^2)((0.5B2)^2 + (F2 + 950)^2)]^{1/2}} \quad (4)$$

Despite the vowel adjustment procedures recommended by Chen, it was not possible to obtain smaller standard deviations (see Table 5). Chen obtained smaller standard deviations primarily because of her subjective peak picking process, as well as a very uniform (and small) speech corpus. Machine-aided peak-finding, particularly over a diverse corpus, is subject to more variability than a fully manual method.

Spectrum-derived parameter	Minimum	Maximum	Mean	Std. Deviation
<i>A1-P1</i> (dB)	-24.09	50.64	13.7942	12.3059
<i>A1</i> Frequency (Hz)	205.00	1265.00	677.3605	201.8381
<i>P1</i> Frequency (Hz)	765.00	955.00	889.3300	61.5894
<i>F2</i> Frequency (Hz)	667.00	3243.00	1697.9605	419.8848

Table 5. Overall means and standard deviations for the parameters obtained with the Chen method

### 2.8.1 Summary of the spectral method

Because a detailed statistical summary of %N was included in 2.7, only a brief summary of the spectral method will be included here (Figure 35; also, compare with **Error!**



**Reference source not found.**) Note that the exact same statistical procedures were applied to both methods.

1. Region,  $F(2,1412)=13.448$ ,  $p<.001$ , and talker sex,  $F(1, 1412)=158.530$ ,  $p<.001$ , were significant, but not their interaction,  $F(2, 1412)=2.526$ ,  $p=.08$ .
2. LM men had significantly higher nasalization levels than both MM,  $F(1, 1412)=7.740$ ,  $p<.025$  and UP,  $F(1, 1412)=19.497$ ,  $p<.001$ , but LM and UP men were not different from each other,  $F(1, 1412)=3.162$ ,  $p=.076$ .
3. LM women had significantly higher nasalization levels than both MM,  $F(1, 1412)=8.200$ ,  $p<.025$  and UP,  $F(1, 1412)=4.144$ ,  $p<.05$ , but LM and UP men were not different from each other,  $F(1, 1412)=.409$ ,  $p=.523$ .<sup>12</sup>
4. Within each region, men and women were significantly different from each other.

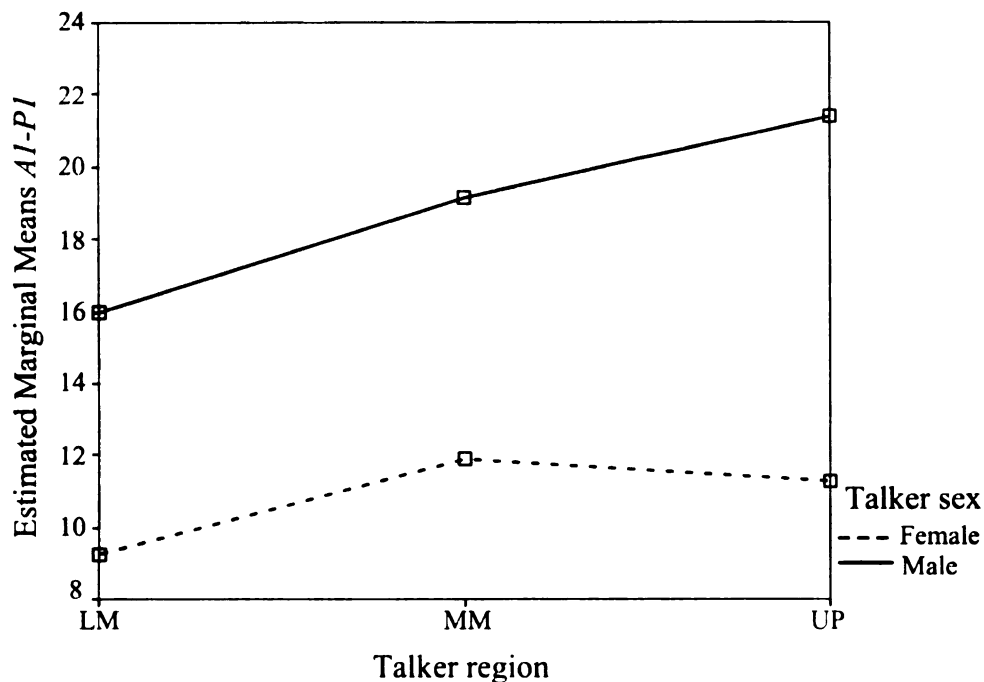


Figure 35. Summary of *A1-P1* results obtained in the spectral method (higher *A1-P1* = lower nasalization)

<sup>12</sup> Note that this result is different than that obtained in the N% analysis, where MM and UP women had significantly different nasalization levels.

## 2.9 NASALIZATION AND NCCS

Having established major patterns of distribution of vowel nasalization in Michigan, it was important to see whether nasalization was in any way related to the on-going vocalic changes that are part of NCCS. As mentioned in 2.5.1, the first spectrographic evidence of increased levels of nasalization in non-nasal environments occurred during standard spectral analysis of vowels from Detroit-area female talkers. Subsequent aerodynamic analysis confirmed high levels of nasalization in this population. Figure 36 shows continuous levels of %N over the course of the words “dad” (dashed line) and “man” (solid line). The levels of nasalization of /æ/ in a non-nasal environment (“dad”) are almost as high as those of /æ/ in a nasal environment (“man”).

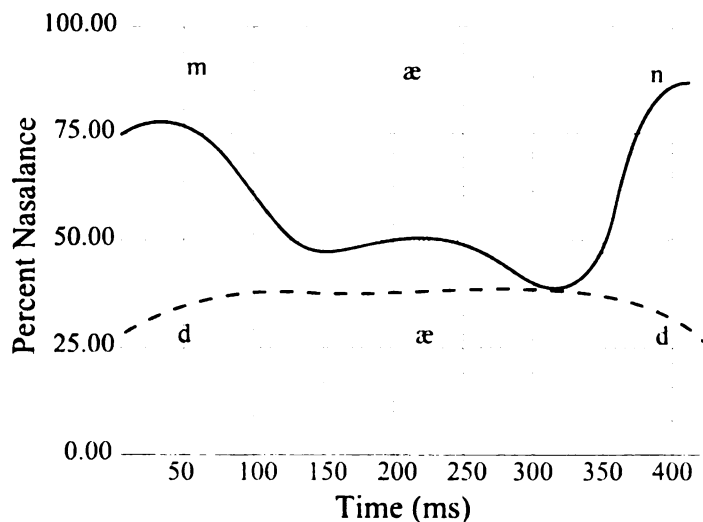


Figure 36. Continuous % Nasalance levels for the words "dad" and "man"

## **2.9.1 The $F2$ of /a/ as an index of talker participation in NCCS**

### **2.9.1.1 Which low-level acoustic feature is the most stable predictor of talker participation in NCCS?**

Finding a good candidate for the marker (or predictor) of talker participation in NCCS is not easy. Ideally, one would need a feature that has the most stable production distribution, as well as perceptual salience. Experimental sociophonetic data on perceptual salience in NCCS is rather difficult to find. However, as will be shown in CHAPTER 3, /a/ has been found to have a robust, generalized perceptual effect. It also happens to be the least controversial from the point of view of both acoustical measurement and sociolinguistic theory, as its movement in NCCS is practically only along the dimension of  $F2$ . Therefore, the  $F2$  of /a/ has been chosen as the best predictor of talker participation in NCCS in the present study.

### **2.9.1.2 A comment on the acoustic analysis of /a/**

It may seem inconsistent to perform acoustic analysis of /a/ in light of possible nasalization issues (see 2.5.1). However, careful analysis of very high-quality speech data with *Akustyk* makes such analysis of /a/ acceptable. *Akustyk* has a number of safeguards against incorrect LPC-derived formant values, though these safeguards do not make the software impervious to serious problems stemming from the analysis of highly attenuated, noisy recordings. First of all, *Akustyk* dynamically estimates optimal LPC filter order based on input sample rate and talker sex. It then performs LPC analysis (either on a static analysis window or pitch-synchronously Figure 37) with the automatically obtained LPC filter order, which it dynamically increments by 2 units

above and 2 below. If it detects possible LPC errors, it displays a warning and allows the investigator to decide which filter order works most reliably before the data are written to the database (see “safety algorithm” in Figure 37). In addition, *Akustyk* can be configured to obtain formant frequencies from narrow-band spectra, and not from LPC. This is particularly useful in the case of nasalized vowels. *Akustyk* first gets LPC-derived values, subject to the investigator’s approval, and then searches for the highest-amplitude harmonics within a user-defined number of FFT bins around LPC-obtained peaks (typically, each bin is 10 or 20 Hz wide). This helps avoid incorrect formant values, such as those obtained from LPC analysis alone (see 2.4.5). In addition to static analysis, *Akustyk* performs interval analysis at 10 ms intervals along the entire duration of the vowel nucleus, which allows the investigator to obtain a detailed account of the spectro-temporal properties of formant trajectories. *Akustyk* employs a formant transition algorithm (see Figure 37) to define formant transitions in acoustic terms. The investigator can choose whether or not to include formant transitions in the interval analysis of formant trajectories (via the “safety algorithm” in Figure 37).

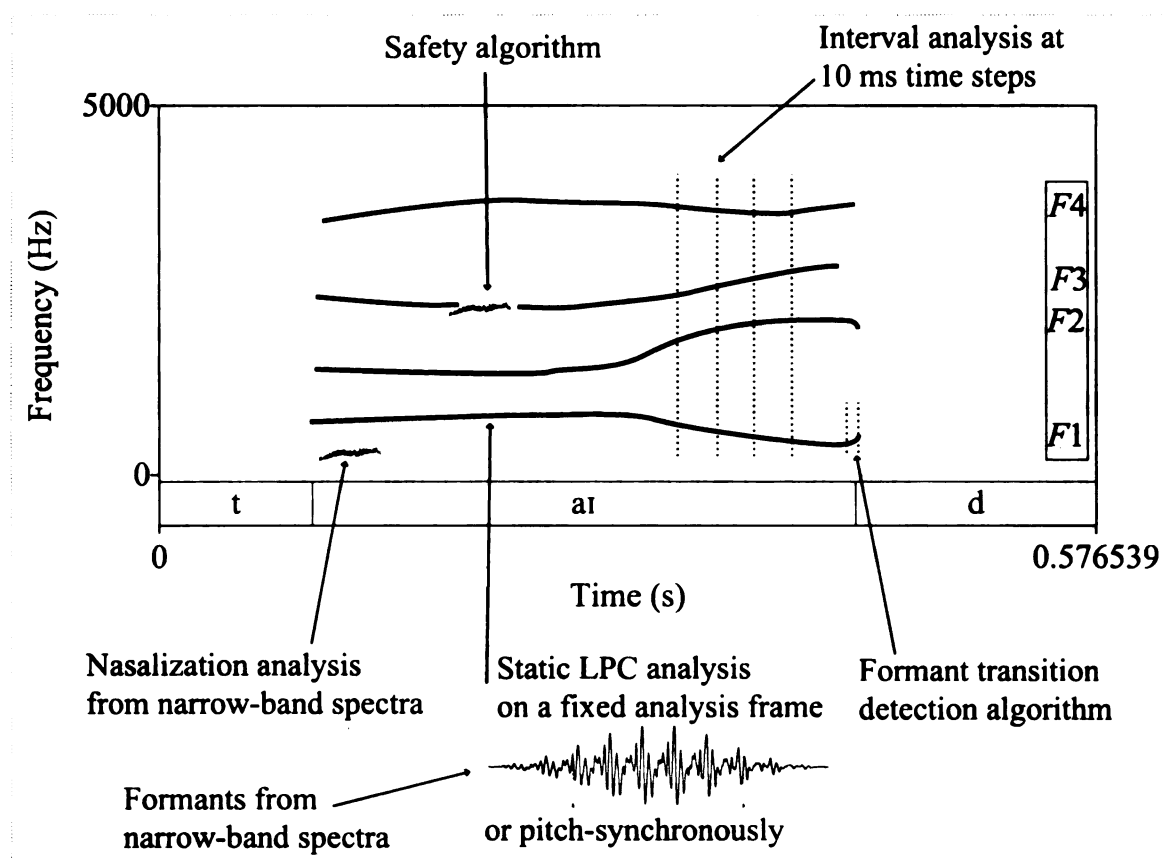


Figure 37. Overview of *Akustyk*'s analysis tools

### 2.9.1.3 Which normalization scheme to choose?

Most modern sociophonetic research employs vowel formant normalization when conducting quantificational analysis among talkers of different sex and/or age (Labov, 2001). Formant normalization is not to be confused with talker normalization (see 1.4.4.2), which is a perceptual phenomenon, though formant normalization aims at achieving a similar effect. It is a computational transformation of raw formant values in Hertz to account for the differences in vocal tract length, most typically, between men and women. There are several normalization schemes with varying scatter reducing power (Adank, 2003). Two of them, Nearey (1977) and Lobanov (1971), use a statistical approach, computing normalization by means of log mean (Nearey) and z-score

(Lobanov) transformations. Nordstrom and Lindblom (1975), on the other hand, normalize by multiplying female formant values by a coefficient derived mathematically from  $F3$  values of male and female talkers in the corpus. Each of these methods has its own advantages – the Lobanov method has the best scatter reducing power. Nearey's formula has been widely used in sociophonetics (Labov, 2001), and Nordstrom and Lindblom's scheme takes into account actual anatomical differences. However, all these normalization schemes are production-oriented. For the present study, it was important to choose a normalization scheme that has real meaning in auditory processing.

#### **2.9.1.4 Why Bark transformation works best**

Vowel nasalization is a direct function of the velopharyngeal port opening. Therefore, one might expect %N not to be related to general anatomical differences among male and female talkers. While it is true that female nasal passages are slightly shorter than those of men, the effect of the possible energy loss due to cavity size differences would probably require a study beyond the scope of this work. For the purposes of this dissertation, one might assume that a 1.5-3cm difference in nasal cavity size would be rather insignificant. Besides, the aerodynamic measurement of %N is derived from the ratio of oral to nasal energy, while the spectral method is based on the ratio of oral and nasal formant amplitudes. Thus, both variables should be, by design, independent of differences in vocal and nasal tract lengths between men and women.

Thus, one is faced with a situation where the dependent variable, %N, is measured on an absolute scale, shared equally by all speakers, and only the  $F2$  of /a/ needs to be normalized for vocal tract length differences. Talkers with the velopharyngeal port open,

will deliver to their listeners' auditory systems a signal whose spectrum contains nasal peaks and zeroes. Somehow, the listeners will have to normalize that signal in order process it correctly. Because vowel nasalization has such a complex hearer-mediated component (see 4.2.2), it was necessary to use a normalization scheme that has a real meaning in auditory processing, which is why the Bark scale transformation (e.g., Zwicker & Terhardt (1980)) was selected to be the most appropriate. Figure 38 shows a cochleagram of the word "back" by a female Detroitter (the same token as that in Figure 27 and Figure 28). Cochleagrams, which combine the features of auditory spectra and spectrograms, use shades of gray to illustrate spectral amplitudes with time, along the x-axis, and frequency, along the y-axis. In the present example, the more common Hertz scale has been replaced with the Bark scale to illustrate the complex auditory properties of the nasalized vowel /æ̃/ as it relates to the notion of "critical distance" (Chistovich, 1985). While "ordinary" spectrograms represent acoustic properties of sounds, cochleagrams constitute their auditory representations. Figure 38 shows two prominent frequencies (dark lines) below 5 Bark (approximately 500 Hz), which indicates their potentially important role in auditory processing. These two peaks result from nasal cavity coupling in ways described in 2.5.1.

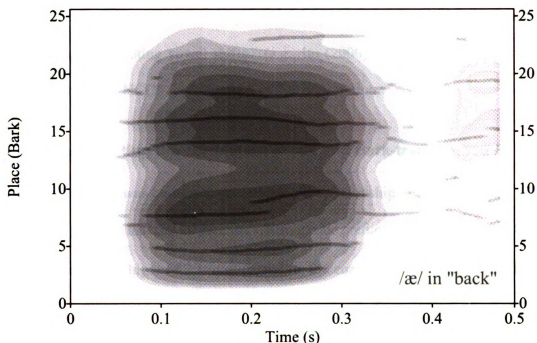


Figure 38. Cochleagram of /æ/ in “back” by a female Detroit

### 2.9.1.5 Bark-transformed $F2$ of /a/

An index of talker participation in NCCS was created, based on Bark-transformed second formant (“fronting”) frequency in Hz. The raw  $F2$  values were extracted from a word list containing 10 instances per subject of /a/ across several different phonetic environments (see Appendix A).  $F2$  extraction and Bark transformation were done with *Akustyk* (Plichta, 2004).

### 2.9.1.6 NCCS and phonetic environments

Chain shifts, including NCCS, are believed to be general in nature, and not specific to any particular phonetic environment, though, of course, some environments may cause formant values to vary quite considerably. This is primarily due to articulatory gestures necessary for the preceding and following consonant, or lack thereof (Stevens & House,



1963). Although sometimes sociophoneticians talk about particular environments as “promoting or demoting vowel shifts,” (e.g., Ito (2000) or Labov (1994)) this is not the meaning in which NCCS is understood in this dissertation.

### **2.9.2 Investigating the correlation of %N and Bark-transformed $F2$**

Linear regression was chosen to investigate the relationship between %N and  $F2$  and to test the hypothesis that the increase in %N might be predicted by the degree of fronting of /a/, as revealed in higher  $F2$  values (based on the corpus described in 2.6.2). The scatter plot in Figure 39 shows the results for 30 subjects where %N is plotted as a function of each subject’s  $F2$  of /a/ (measured in Bark). The results indicate that the two variables are linearly related such that as  $F2$  increases, so does %N. The scatter plot of Bark-transformed  $F2$  and mean %N follows the regression line in a regular fashion.

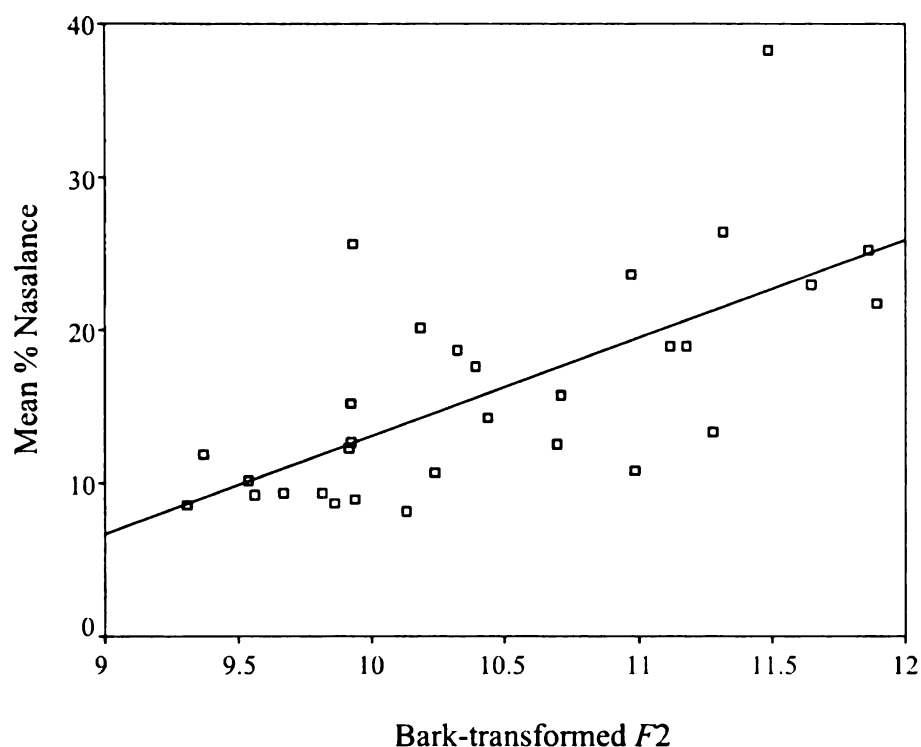


Figure 39. Scatter plot of Bark-transformed  $F2$  and Mean % N fitted around the regression line

### 2.9.2.1 Regression results

The regression equation for predicting the overall %N scores was:  $\text{Predicted \%N} = 6.39 \text{ Bark-transformed } F2 - 50.811$ . The correlation between Bark-transformed  $F2$  and %N was .681,  $t(28)=4.9$ ,  $p<.001$ . Bark-transformed  $F2$  also explained a significant proportion of variance in %N,  $R^2 = .463$ .

## 2.9.3 Summary

### 2.9.3.1 Summary point #1

There is evidence to suggest that traditional signal acquisition and analysis methods may have lead researchers to imprecise claims about some NCCS vowels, particularly about the raising of /æ/ and the lowering of /ε/.

### **2.9.3.2 Summary point # 2**

There is evidence that some Michigan talkers pronounce vowels in non-nasal environments with spectra containing components of nasal energy.

### **2.9.3.3 Summary point # 3**

It has been established that %N is distributed in Michigan along the sociolinguistic isoglosses of sex and region. Women have higher levels of %N both across and within the three regions (LM, MM, and UP).

### **2.9.3.4 Summary point # 4**

The %N distribution among men is different. LM men have higher levels of %N than both MM and UP, but there is no significant difference between MM and UP males with regard to %N.

### **2.9.3.5 Summary point # 5**

%N is correlated with Bark-transformed  $F2$  in the Michigan sample. Since  $F2$  is a strong predictor of NCCS participation, it is likely that %N is distributed along the lines similar to those of NCCS.

### **2.9.3.6 Summary point # 6**

The results obtained in the machine-aided spectral analysis of nasalization yield additional support for the fact that region LM (most influenced by NCCS) is ahead of both MM and UP in levels of vowel nasalization.

## **CHAPTER 3**

### **PERCEPTIONS OF /a/-FRONTING**

#### **3.1 INTRODUCTION**

##### **3.1.1 Introduction to talker normalization**

###### **3.1.1.1 What is talker normalization?**

The speech signals that a listener must process are extremely variable (Ohala, 1981). The sources of this variability are of several different kinds. One is between-talker variability in vocal tract length (VTL). VTL varies primarily across talker sex (men vs. women) and age (children vs. adults), and it directly affects formant frequencies of all vowels produced by a talker (Stevens, 1998). For example, in Figure 40, the resonant frequencies (formants) of the vowel /æ/, as in “back,” produced by males and females from the same speech community may vary due to VTL differences by as much as 300 Hz along  $F_1$  and  $F_2$ . Three hundred Hz is sufficient to produce confusions among vowels in theory (Peterson & Barney, 1952), yet listeners generally have little trouble with correct vowel identification. They apparently make a perceptual “correction” for talker differences, a phenomenon referred to as “talker normalization” (Nearey, 1989). Figure 40 shows an  $F_1/F_2$  plot of mean non-normalized male and female formant values from the Peterson and Barney corpus (1952). Note that most of the male and female vowels are distributed along very similar lines, i.e., they occupy similar relative positions in the vowel quadrilateral. The female formant values are generally higher in frequency due to VTL differences, as shorter vocal tracts produce higher frequency formants. Other common types of variability for which listeners must also perceptually adjust include variation in speaking rate and discourse frame (Lindblom, 1963). Vowels generally become reduced (more schwa-like) with increase in speaking rate and decrease in syllable stress. Finally,

there is notable vowel variability due to differences in consonantal context (Stevens & House, 1963).

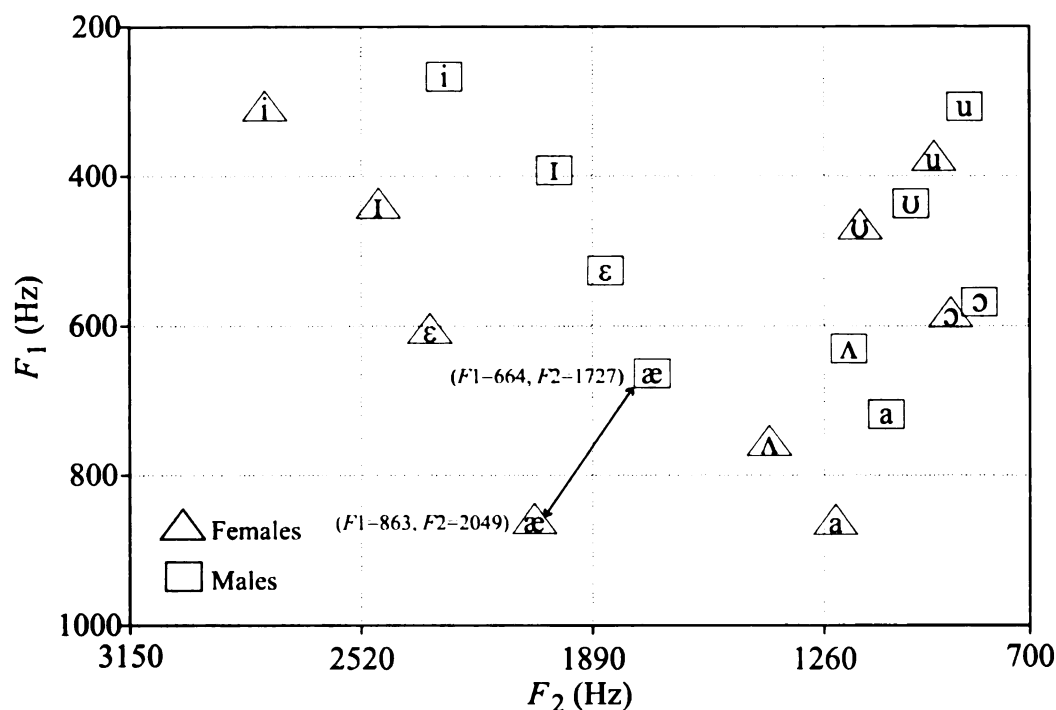


Figure 40. Differences in  $F_1$  and  $F_2$  due to VTL variability between men and women (from the Peterson and Barney corpus (1952))

### 3.1.1.2 Intrinsic normalization

Proponents of intrinsic normalization theory argue that there is information within the spectral content of the vowel itself to support a listener's normalization computation. Two of the most convincing studies in intrinsic normalization are those by Miller (1989) and Syrdal and Gopal (1986). Miller demonstrated that monophthongal vowels of American English can be represented as clustered in perceptual target zones within a three-dimensional ( $F_3$ - $F_2$ ,  $F_2$ - $F_1$ , and  $F_1$ -SR<sup>13</sup>) auditory perceptual space. Miller

<sup>13</sup> "SR" is "Sensory Reference" – a reference model proposed by Miller.

examined a number of speech corpora including the Peterson and Barney corpus (1952) in this way. Syrdal and Gopal used a Bark transformation to devise a two-dimensional model of vowel recognition based primarily on the perception of critical distance, in Bark (Chistovich, 1985). The evidence for intrinsic normalization came from successful discriminant analysis of vowels in this derived vowel space.

A limitation of these studies is that they did not include sources of sociolinguistic variability as it relates to vowels. Both studies used the Peterson and Barney corpus, which contains a well-documented, but limited database of formant values from a select set of monosyllabic words; all produced in /hVd/ context. Very little is known about the speakers' dialect history. The other corpora used by Miller are also void of documented sociolinguistic variability.

To illustrate the potentially important role of sociolinguistic (dialectal) variation, a Bark analysis, similar to that of Syrdal and Gopal (1986) was applied here to a corpus of NCCS vowels and to the Peterson and Barney corpus. Figure 41 shows a two-dimensional plot of the Peterson and Barney data delimited by  $F3-F2$  and  $F1-F0$  in the discriminant plane. It can be seen that each vowel occupies its own unique (non-overlapping) spot in the discriminant space (delimited by the ellipses), which suggests that, with intrinsic normalization, there should, in principle, be very little vowel confusion.

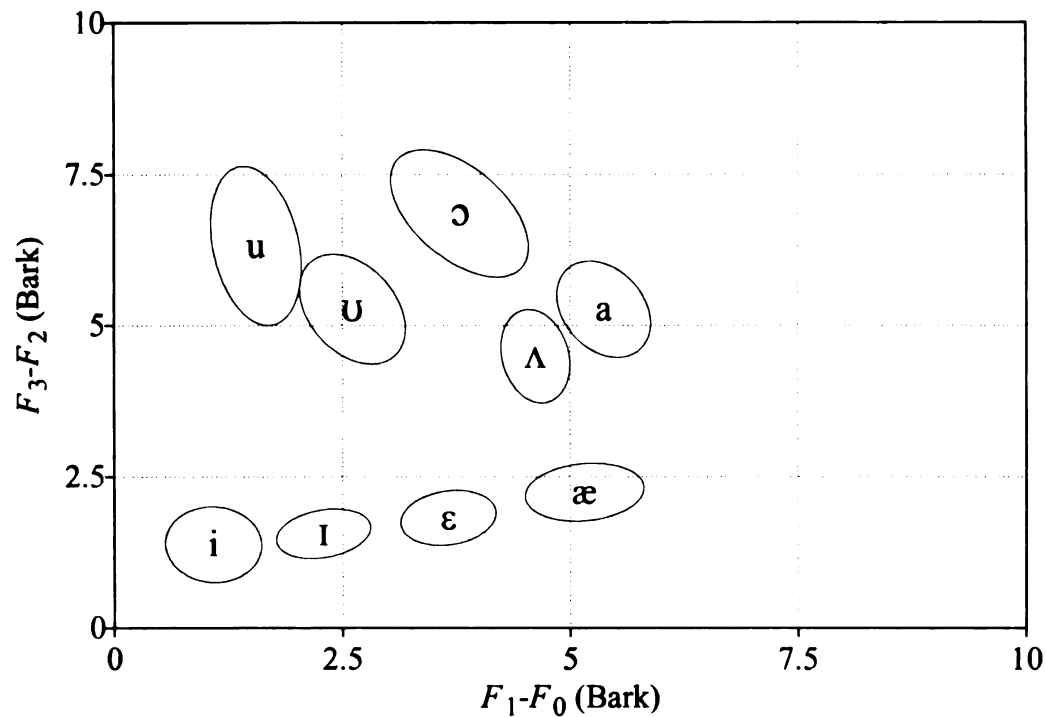


Figure 41. Bark-transformed, discriminant plane of formant values from the Peterson and Barney corpus

Figure 42 shows a Bark-transformed discriminant plane of formant values from a corpus of 26 adult NCCS speakers.<sup>14</sup> There is substantial overlap among /æ/, /ε/, /ɔ/, and /Λ/ in this corpus, which suggests that confusion would be more likely to occur in communities where NCCS is in progress. These results also suggest that in such communities the success of intrinsic normalization strategies, such as that of Syrdal and Gopal, is less likely.

<sup>14</sup> The Peterson and Barney corpus is often praised by speech scientists and speech engineers for its rigorous signal acquisition methods. This reputation is well deserved. However, modern technology, particularly, 24-bit digital audio recorders and flat-response microphones, often make today's signal acquisition methods even more reliable than those of Peterson and Barney.

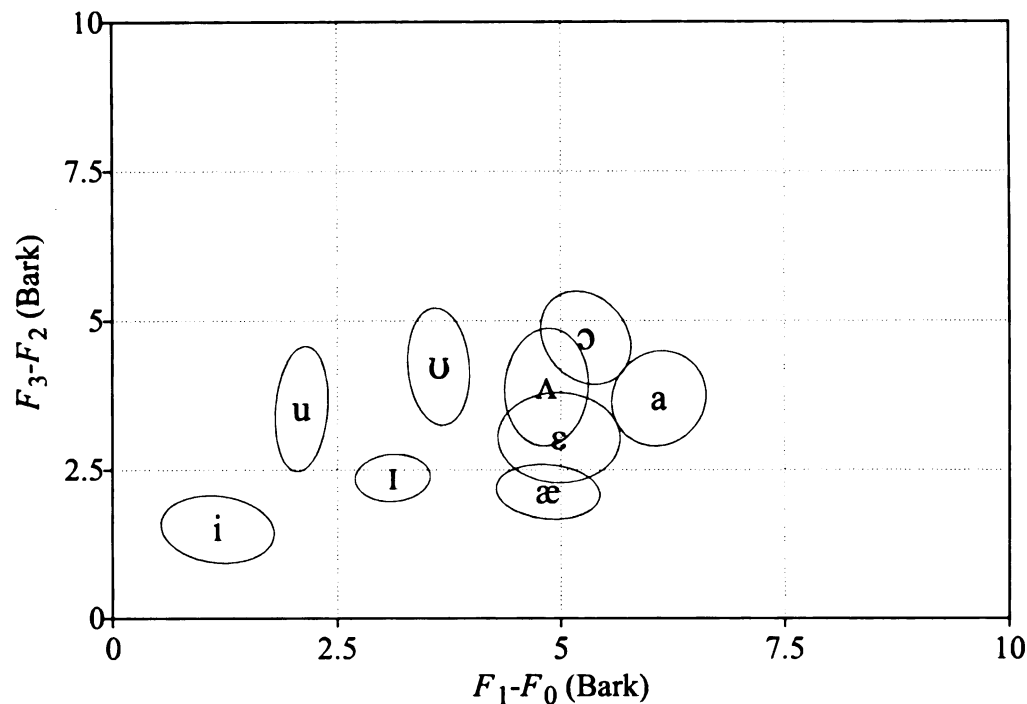


Figure 42. Bark-transformed, discriminant plane of formant values from a corpus of 26 adult NCCS speakers

### 3.1.1.3 Extrinsic normalization

Proponents of extrinsic talker normalization argue that talker normalization requires a perceptual frame of reference that is established based on information obtained from sources beyond the vowel itself. It might, for example, come from information obtained elsewhere in the utterance, or from extralinguistic sources, such as visual information about the talker (e.g., Massaro, (1998)). The subsequent sections of this chapter will be devoted to a version of extrinsic talker normalization that can be termed “sociophonetic talker normalization.”

### 3.1.1.4 Sociophonetic talker normalization

Sociolinguists are interested in articulatory-acoustic variability due to socially-constructed behavior. These low-level phonetic variants are essential to our



understanding of language variation, as they help us identify social forces behind language change.

Sociophonetic talker normalization focuses on the integration of sociolinguistic (or dialectal) features both in the top-down and bottom-up processing of speech (see 1.4.3). Sociophoneticians are interested in testing the hypothesis that the variable acoustic signal carries a wealth of sociolinguistic information, such as talker age, gender, socio-economic status, educational background, and so forth. Over time, repeated exposure to this information may result in the development of an abstract representation of idiolect and dialect. Folk linguistic studies, such as those mentioned in CHAPTER 1, have demonstrated this fact fairly convincingly.

#### **3.1.1.5 Information conveyed by vowels**

Ladefoged and Broadbent (1957) demonstrated that linguistic information conveyed by a vowel does not depend on the absolute values of its formant frequencies, but on the relationship between the formant frequencies of that vowel and the formant frequencies of other vowels produced by that speaker. Being one of the first studies to use a speech synthesizer, it provided a compelling account of talker normalization vis-à-vis talker physiology. By shifting formant values of the entire vocalic system of a synthetic precursor phrase by a constant value, Ladefoged and Broadbent obtained different perceptions of specific  $F1/F2$  vowel patterns that followed the precursors.

## **3.2 THE STUDY**

### **3.2.1 Motivation for the study**

Ladefoged and Broadbent (1957) pointed out three types of information conveyed by vowels: linguistic information, sociolinguistic information, and personal information. While their experimental study focused on the relationship between personal information and vowel perception, they made two interesting observations following their experiment:

1. "There is tentative evidence that subjects belonging to different sociolinguistic groups gave different responses to some of the test material" (p. 103).
2. "It seems at least possible that both linguistic and sociolinguistic information conveyed by vowels depends largely on the relative positions of their formants" (p. 103).

Surprisingly, the provoking ideas of this 1957 study have not been given much attention within the sociophonetic community. The present study was designed as a first attempt to answer Ladefoged and Broadbent's unanswered questions regarding sociolinguistics. The specific research questions were formulated as follows:

1. Does talker-dependent, sociolinguistic information influence speech perception?
2. Do hearer-dependent, sociolinguistic factors influence speech perception?
3. What is the nature of talker normalization involved in this process?
4. Is /a/-fronting perceptually salient among NCCS speakers?

### **3.2.2 Lower Michigan and the Upper Peninsula**

Lower Michigan (LM) and the Upper Peninsula (UP) (Figure 43) are two different dialectal regions of Michigan. They differ in other ways as well. Ishpeming in the Upper

Peninsula, for example, is a small town in the vicinity of the Marquette Iron Range. It is a working-class town and it is known for iron-mining activities, lumbering, marble quarrying, and winter sports. The population of the Ishpeming-Marquette area is mostly of Scandinavian origin. Detroit in Lower Michigan, on the other hand, is a large metropolitan area. It is known for being a center of American automotive industry. Detroit is a very dynamic, ethnically, and linguistically diverse city. The European American middle-class and upper-middle class majority have been reported to speak a dialect of English influenced by NCCS (Labov et al., 1997).



Figure 43. Lower Michigan and the Upper Peninsula

Linguistically, the dialect of English spoken in the Upper Peninsula is distinct from that of Lower Michigan. It is under Canadian influence (Canadian raising) and it rarely shows elements of NCCS. The standard *F1/F2* vowel chart below (Figure 44) contains normalized (Nordstrom & Lindblom, 1975) mean formant data extracted from recorded speech samples from the UP and LM men ( $n=10$ ) and women ( $n=9$ ) participating in the

present study. Most of the differences are in the vowels /a/, /æ/, and /ε/, which is not surprising as those three vowels are active participants in NCCS. Table 6 contains normalized mean  $F_1$  and  $F_2$  values for the two groups obtained from the word list in APPENDIX F.

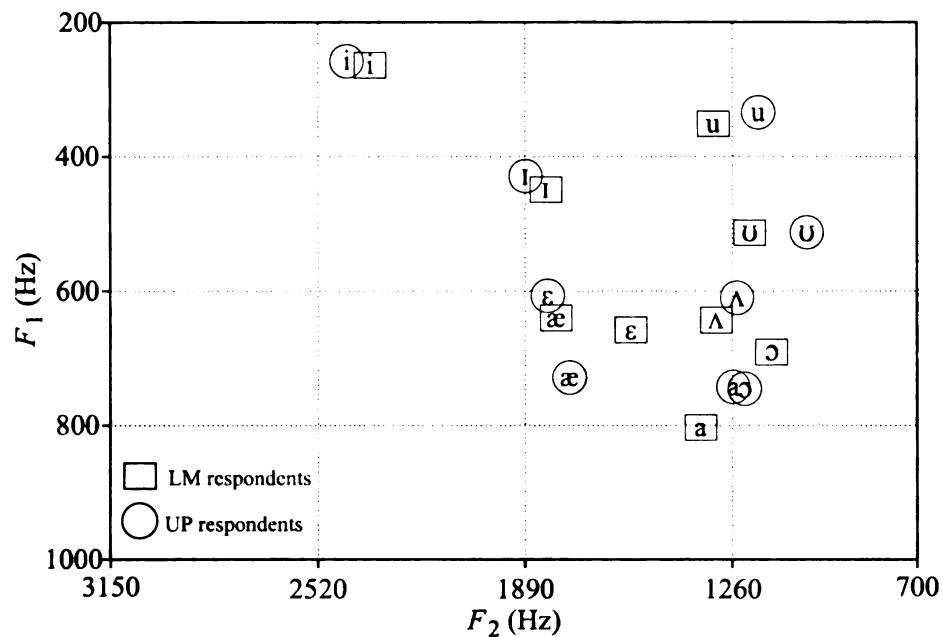


Figure 44. Normalized, mean formant values of LM and UP participants in the study

Vowel	LM respondents		UP respondents	
	<i>F1</i>	<i>F2</i>	<i>F1</i>	<i>F2</i>
/i/	264	2361	258	2432
/ɪ/	449	1826	429	1888
/ɛ/	658	1569	607	1822
/æ/	640	1796	728	1755
/a/	803	1357	742	1258
/ɔ/	690	1142	745	1223
/ʊ/	513	1207	513	1034
/u/	351	1318	334	1182
/ʌ/	643	1310	610	820

Table 6. LM and UP respondents' normalized, mean *F1* and *F2* values in Hertz

### 3.2.3 Talker normalization across the LM and UP dialects

If listeners make a perceptual adjustment for dialectal differences in a talker's production of vowels, it might be expected that the same vowel sound input would be interpreted differently depending on dialectal context. This perceptual effect might also depend on the listener's dialect history. These predictions were tested in the present study.

### 3.2.4 The subjects

The subjects were recruited from the Detroit area of Lower Michigan (5 men, 5 women, ages 19-30) and from the Ishpeming area of the Upper Peninsula (5 men, 4 women, ages 19-34). All of the subjects were of European American descent. The subjects had to have been born and raised in their respective region and to have never left their region for

more than a year. They had to be native speakers of English. Figure 44 shows their normalized, mean vowel systems within the  $F1/F2$  space. Similar to the study in described in 2.6.2, the subjects were selected from among groups that would, in theory, be likely participants in NCCS.

### **3.2.5 The stimuli**

#### **3.2.5.1 Speech synthesis**

As mentioned in 1.4.4.3, synthetic speech has often been employed in speech perception research. There exist several different types of speech synthesis, out of which three are the most common: parametric speech synthesis, synthesis from voice samples, and LPC analysis/re-synthesis. Parametric synthesis creates speech-like digital samples from numerical input to common speech production parameters, such as oral constriction, degree of breathiness, voicing, and others. This type of synthesis is the most difficult to use in large-scale, commercial projects, such as automated phone services, but it is very effective in small-scale empirical research as it allows precise control of the signal. Speech synthesis from voice samples, or concatenative synthesis, (synthesis from digitally pre-recorded diphones or syllables), on the other hand, does not lend itself to detailed perceptual research, but has been very successful in creating realistic-sounding samples, particularly in prosodic terms (Rodman, 1999). It is also the most commonly used speech synthesis type in commercial text-to-speech (TTS) applications. The third method, LPC analysis/re-synthesis (see the experiment described in 1.4.2.3) produces very realistic-sounding samples from real speech and allows a certain degree of manipulation, particularly in terms of formant frequencies, bandwidths, and amplitudes. There are other, similar methods such as PSOLA (Pitch-synchronous Overlap and Add),

which allows the re-synthesis of pitch and duration. LPC and PSOLA can be used in tandem. Since both LPC and PSOLA crucially rely on voiced samples, they are, therefore, most effective with voiced speech.

### **3.2.5.2 The hVt and sVk continuum**

The choice of speech synthesis methodology was crucial to the success of this study. Initially, the LPC analysis/re-synthesis method was considered but due to its reliance on real speech samples it had to be eliminated. It was important to create a stimulus that contained a very small, highly controlled set of acoustic parameters. LPC-based re-synthesis produces samples that contain too much speaker-specific, and hence, uncontrollable, spectral detail. Therefore, a parametric synthesis method was chosen as the primary tool in the production of target word stimuli. A parametric synthesizer allows the researcher to produce realistic-sounding samples, while controlling all of the acoustic parameters involved.

A 7-step /a/~/æ/ (hVt and sVk)<sup>15</sup> continuum was synthesized with the Hlsyn parametric synthesizer (Sensimetrics, 1997), which is based on the Klatt synthesizer (1990). Since the study focused on the fronting of /a/, the synthetic vowel continuum was created along *F2*, with appropriate formant transition adjustments for the preceding and following consonants. The continuum was based on real formant data measured from two young, middle-class males – one from Lower Michigan (Talker LM) and one from The Upper

---

<sup>15</sup> hVk and sVk will be henceforth used to mean “h + vowel + t” and “s + vowel + k” respectively.

Peninsula (Talker UP) (see also 3.2.5.3 for detailed information about the two talkers).

The synthesized vowels had the following properties:

1.  $F_0$  – falling from 120 to 100 Hz.
2.  $F_1$  – fixed at 750 Hz (mean value of Talker UP and LM's /a/ and /æ/).
3.  $F_2$  – 1243-1441 Hz (33 Hz intervals)<sup>16</sup>.
4.  $F_3$  – fixed at 2500 Hz, with formant transition adjustments.

A speech continuum is a common tool in speech perception experiments (see, for example, 1.4.2.3 and 1.4.4.3). Figure 45 shows first and second formant tracks of each of the seven steps of the /a~/æ/ continuum in “sock~sack” used in the study. As can be seen,  $F_1$  is kept constant, while  $F_2$  increases incrementally (by 33 Hz) from 1243 Hz to 1441 Hz (along the top of Figure 45). At some point in the continuum, the vowel moves across the contrastive (phonemic) categories of /a/ and /æ/ (indicated by the arrow along the top). The exact point at which this occurs was the main focus of this investigation. It was surmised, based, for example, on the Ladefoged and Broadbent (1957) study, that the category cross-over point (see 3.2.7.2 for a detailed discussion) might be different depending on the stimuli involved (such as the dialect of the speaker of the precursor phrases in 3.2.5.3), though the exact nature of this relationship was unknown. Also, note that the formant tracks exhibit shapes characteristic of formant transitions coming out of

---

<sup>16</sup> This continuum was initially synthesized in 11-steps. A pilot study helped eliminate 2 lowest and 2 highest of the original steps because responses had plateaued prior to these endpoints of the sequence.



the preceding consonant /s/ and into the following consonant /k/. This was done in order to obtain a naturalistic sample.

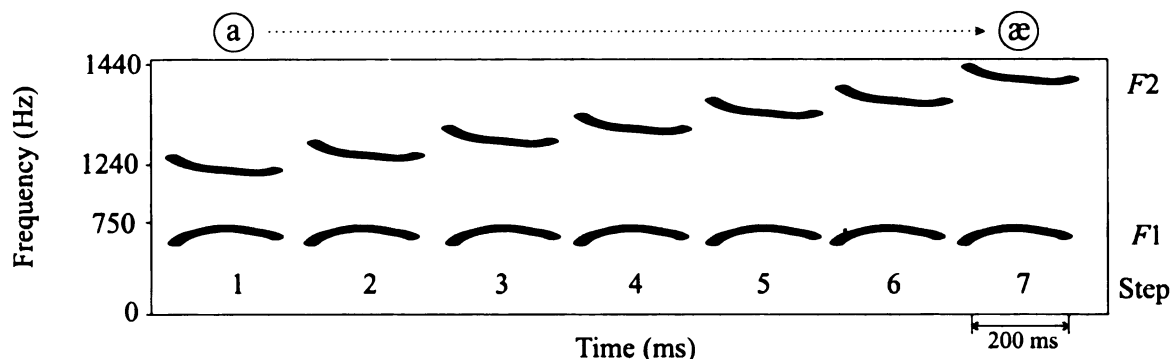


Figure 45. *F1* and *F2* tracks of the 7-step continuum of “sock-sack” used in the experiment

Figure 46 shows a spectrographic contrast between steps 1 and 7 in /sVk/ syllable context, while Figure 47 contains LPC spectra obtained from the same pair of vowels. Note that the primary spectral differences between these two steps are mostly in the absolute values of *F2* in Hertz. As can be seen both in the spectrograms and in the LPC plots, the synthetic samples have speech-like properties. The spectrograms, for example, show a fair amount of frequency perturbation characteristic of speech while the LPC spectra show realistic-looking spectral envelopes with well-defined formants, narrow bandwidths, and properly attenuated formant amplitudes.

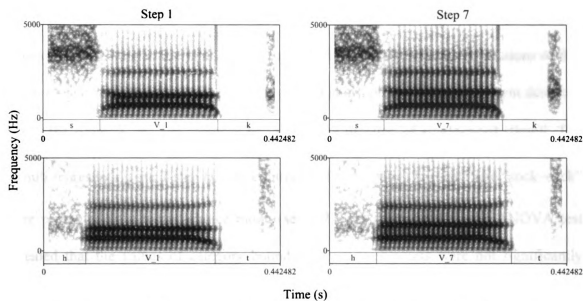


Figure 46. Spectrographic images of the first step and last step of the /a/-/æ/ continuum

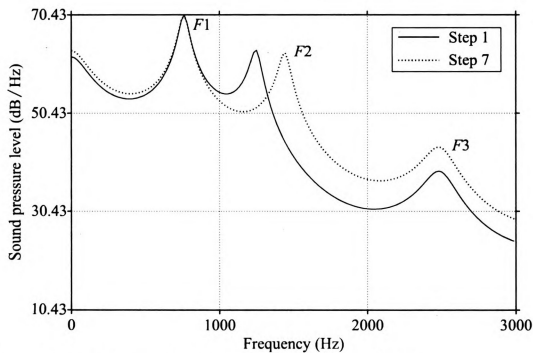


Figure 47. LPC spectra of the first and last step of the continuum

### 3.2.5.3 Preliminary experiment

Prior to participating in the main experiment described in the subsequent sections of this chapter, each subject was asked to participate in a short, preliminary experiment designed to test the subjects' /a~/æ/ category boundary as a function of single-word stimuli. The stimuli representing the 7-step /a~/æ/ continuum for the pairs “hot~hat” and “sock~sack” were presented in a forced-choice mode (see 3.2.6 for more details). An ANOVA test revealed that the values of category boundary cross-over points were not significantly different for the two groups of respondents ( $F(1,16)=.06, p=.81$ ). This indicates that even though there were phonetic differences between their vowel systems (/a/, /æ/, and /ɛ/, in particular), both groups shared a similar, general representation of the /a~/æ/ category boundary. It was, therefore, of a great deal of interest to see whether this boundary would shift if the target words were preceded by precursor phrases spoken by talkers representing different dialect regions.

### 3.2.5.4 Precursor phrases

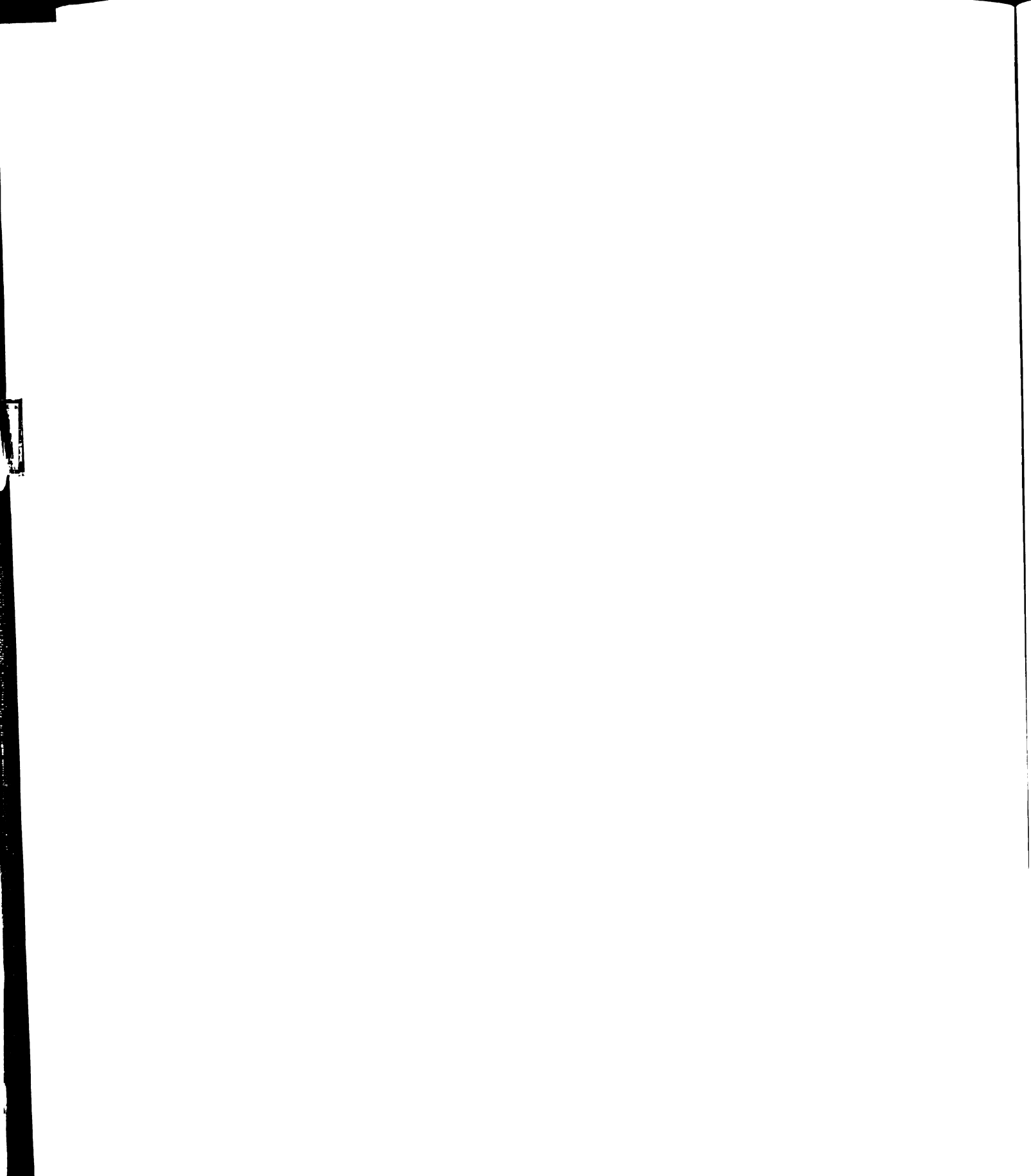
Two sets of four semantically neutral phrases were recorded with a close-talking, flat-response microphone (Sennheiser HMD25-1) and a Tascam DA-P1 DAT recorder at 48,000/16 bit, one set by Talker LM and one by Talker UP. The precursor phrases were designed to contain a broad sampling of the talkers' vowels and to include specific exemplars of /a/ and /æ/. At the end of each phrase, the talkers were asked to say the syllable “uh” in order to complete the phrase prosodically. There precursors were as follows:

1. Bob was positive that he heard his wife, Shannon, say "uh."
2. Cathy's card was blue and said: "pot", while Mary's was black and said: "uh."
3. The key to winning the game of boggle is to know lots of short words like "uh."
4. It turned out that the most common response to question thirty two on last week's test was: "uh."

Talkers LM and UP were of similar age and  $F_0$ . They were selected carefully to minimize possible VTL normalization effects. Their vowel systems, as shown in Figure 48, were similar, except for the NCCS-influenced vowels /a/, /æ/, and /ɛ/<sup>17</sup>. As evident from Figure 48 and Table 7, Talker LM had a vowel system showing advanced stages of NCCS in ways consistent with Labov (1991), except for /æ/-raising. The arrows point from Talker UP to Talker LM to illustrate the NCCS movement of /a/-fronting (+366 Hz), /æ/-fronting (+380 Hz), and /ɛ/-lowering (+207 Hz). These are substantial, but not unrealistic, differences. Also, note that the rest of the vowels in the chart are in close proximity to one another, which, most likely, indicates both similar dialectal features and similar VTL properties of the two speakers.

---

<sup>17</sup> Formant data were obtained from the wordlist in APPENDIX F



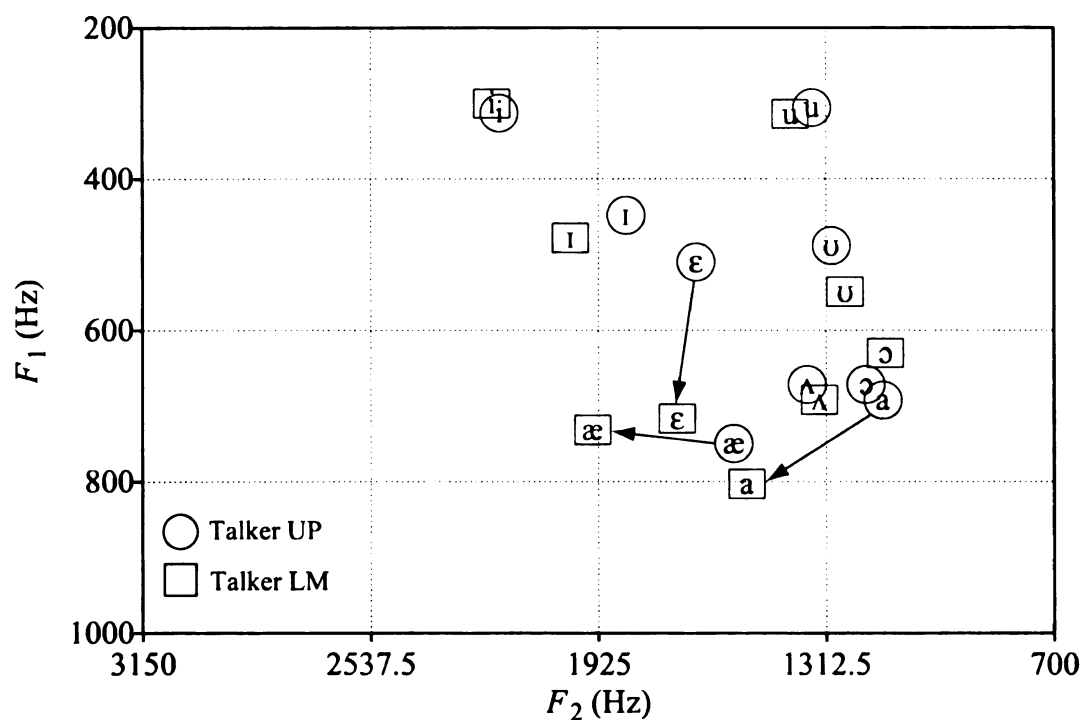


Figure 48. Vowel systems of talkers LM and UP

Vowel	Talker LM		Talker UP	
	$F_1$	$F_2$	$F_1$	$F_2$
/i/	300	2210	312	2190
/ɪ/	478	1999	448	1850
/ɛ/	717	1712	510	1662
/æ/	732	1940	750	1560
/a/	803	1525	692	1159
/ɔ/	630	1152	671	1205
/u/	549	1261	488	1298
/ʊ/	314	1407	305	1349
/ʌ/	692	1330	671	1363

Table 7. Talker LM's and UP's mean  $F_1$  and  $F_2$  values in Hertz

### 3.2.5.5 Finalized stimuli

Finalized stimuli consisted of a carrier phrase with one of the target word variants at the end (Figure 49). The original recordings of Talkers LM and UP were downsampled to 11,025 Hz to match the sample rate of the synthetic target word. The real and synthetic speech samples were RMS (root-mean-square) peak-level normalized to ensure uniform levels throughout the stimuli. The samples were then merged and checked for potential problems, such as digital clicks. It was crucial that the synthetic and real speech samples be as close to each other in sound quality as possible. This is why special care was taken to acquire flat-response recordings with as little extraneous noise as possible. Figure 50 shows a spectrogram of a fragment of precursor phrase 1 (see 3.2.5.3). The formants (dark bands) appear visibly strong and have narrow bandwidths. There is no evidence of unwanted noise or spectral bias in this recording.

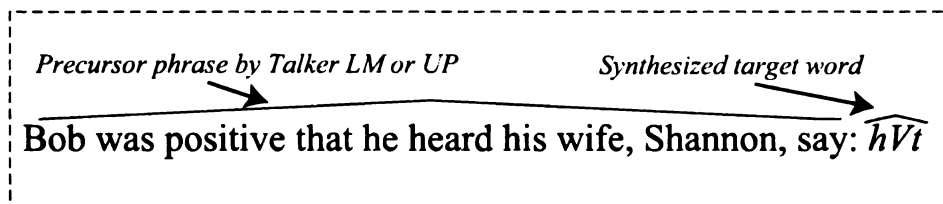


Figure 49 Finalized stimulus: a UP or LM precursor phrase with a synthesized target hVt or sVk word at the end

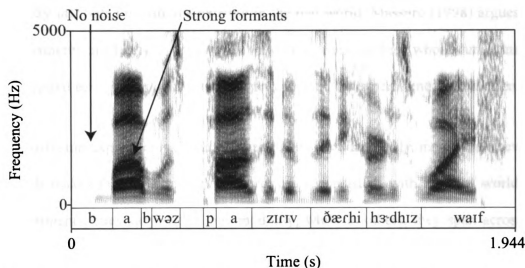


Figure 50. Spectrogram of a carrier phrase fragment by Talker UP

### 3.2.6 The experiment

#### 3.2.6.1 Forced-choice experiments – single and multi-factor designs

Many of the early categorical perception studies (see detailed discussion in 1.4.4.3) were based on a single-factor, forced-choice design. In these types of experimental conditions, all variables, but one are eliminated or made neutral, and only the target variable (such as voice onset time, or VOT) is systematically changed across the stimuli (e.g., Abramson & Lisker (1968)). Such stimuli, typically on an  $n$ -step continuum, are presented to listeners randomly and, at each trial, the subjects are asked to choose one of the two possible contrastive categories (e.g., /p/ or /b/). The most important advantage of a single-factor design is that the researcher is able to control the input variable very well. At each trial, the experimenter knows which specific acoustic parameter triggered which specific perceptual response. However, single-factor experimental conditions offer a limited model of the real world, where multiple factors vary simultaneously. Therefore, even if respondents do exhibit significant behavioral patterns, these patterns often cannot be



directly attributed to patterns existing in the real world. Massaro (1998) argues that such experiments are likely to investigate functional cues, or cues whose functional value in perceptual ecology is not as important as single-factor experiments would often indicate.

Multi-factor experiments, on the other hand, simultaneously manipulate several cues, which makes the experimental situation more consistent with the real world. In some experiments, cues vary within one modality, while in others, they span across different modalities (see, for example, 3.1.1.3). Another important advantage of multi-factor experimental design is that the investigation of need deals with a much larger data set with more inter-dependent variables. It can thus be argued that such large and diverse data sets provide a better empirical coverage of human behavior, thus, potentially, allowing more robust theoretical claims (Massaro, 1998).

### **3.2.6.2 The /a/-fronting study as a multi-factor experiment**

The /a/-fronting experiment was of a multi-factor, forced-choice design. While all cues varied only within the auditory modality, they simultaneously varied across:

1. target word (2 levels – “hot~hat”, “sock~sack”)
2. target word  $F2$  (7 levels – see the sVk continuum in Figure 45)
3. precursor phrase dialect (2 levels – Talker LM and UP)
4. precursor phrase type (4 levels – see the precursors phrases in 3.2.5.3)

If one assumes that the speech signal contains a great deal of simultaneously varying auditory cues and that it needs to be normalized in order to be processed effectively (e.g.,

Ohala (1981)), then the experiment provided listeners with at least 4 types of such acoustic-phonetic cues and obliged them to employ a normalization process more complex than that present in single-factor categorical perception experiments. One could, therefore, argue that, at least from the theoretical point of view, the experimental design of the present study would allow obtaining empirically and behaviorally viable findings.

### **3.2.6.3 Experiment delivery on a laptop computer**

Variationist sociolinguistic research has been dominated by fieldwork (Chambers, 1995). Traditionally, sociolinguists prefer to collect their data at a place of the subject's own choosing, such as their home or place of employment. The argument of observer's paradox (Labov, 2001) demands that fieldworkers not create an atmosphere in which the authenticity and naturalness of the language sample would be compromised. Still, some believe that certain micro-level phonetic dialectal features are beyond the level of talkers' conscious interpretation and evaluation and should, thus, not be subject to variation vis-à-vis observer's paradox (Labov, 2001). Speech science research, on the other hand, has not been preoccupied with observer's paradox to the same extent, and most experimental research has taken place in speech laboratories where the subjects' awareness of experimental conditions is emphasized but where experimental conditions can be controlled very well. While designing a sociophonetic speech perception study, one had to observe the main principles of sociolinguistic field research while providing controlled and rigorous experimental conditions. Therefore, a compromise was reached whereby the experiment was to be run on a modern laptop computer at a quiet place of the subject's choice. Some respondents chose to have the experiment run at their own house, while others chose a public library quiet study room or a vacant conference room.

The advantages of running a speech perception experiment on a modern laptop computer are undeniable. First of all, it was possible to move the experiment out of the speech laboratory, which enabled the investigator to travel to Detroit, MI and Ishpeming, MI in search of eligible respondents. Also, the 16-bit audio capability of a modern laptop computer is close in quality to that of older speech laboratory stationary workstations. While it is true that many built-in computer audio chips produce certain levels of noise from electrostatic interference (Pohlmann, 2000), this noise is audible only at volume levels far greater than those used in the study. Volume optimization was an integral part of the experiment's design. Initially, it was believed that the relatively low level output of the laptop's D/A converter would not be adequate to drive high-quality, 60  $\Omega$  headphones. A portable Rolls 4-channel stereo headphone amplifier was used to experiment with different volume levels. However, it was finally determined that the built-in audio chip provided sufficiently high volume amplification for optimal listening experience.

Perhaps the most serious disadvantage of delivering the experiment on a laptop computer was the fact that touchpad keys had to be used as a response input device. Computer keyboards have a resolution of approximately 20-35 ms, while specially designed response pads can achieve latencies of approximately 1 ms. The laptop computer did not have a functioning serial port, which prevented the investigator from using one of the commercially available response pads. In the near future, USB response pads are expected to be manufactured, which will make them compatible with most modern laptop computers.

At the start of the experiment, the subjects were seated in front of the screen wearing closed, flat-response headphones (Koss R80), necessary both for high quality audio reproduction and for environmental noise attenuation. When presented with a stimulus, the respondent's task was to decide whether the target word sounded more like /a/ (as in "sock") or /æ/ (as in "sack") and to report their choice by pressing the appropriate button on the computer's touchpad. At each step of the experiment, the subjects were automatically given instructions and prompts on the computer screen. The randomized stimuli were presented in eight blocks of 56 trials and, within a block, each stimulus was repeated four times. The blocks were divided into two parts with a 30 minute break in between. Each subject responded to a total of 448 trials.

### **3.2.7 Analysis and results**

The results of each trial were written to a database in real time. Each trial was coded for reaction time<sup>18</sup>, dialect of precursor phrase, word, phrase, etc. The data were then dynamically merged with demographic information about the subjects into a comma delimited non-proprietary ASCII data file compatible with statistical analysis software and appropriate for long-term preservation. The data were cross-tabulated for each subject to obtain information about individual differences. A psychometric function was plotted for each subject and for each subject group.

---

<sup>18</sup> Only accuracy data were analyzed for the present study.

### 3.2.7.1 Psychometric functions

The next step in the analysis was to sub-divide the overall perceptual data set in a principled way and make a psychometric comparison of the parts. The 7-step stimulus sequence moved incrementally in physically equal steps of 33 Hz. Psychometric functions such as those in Figure 51 and Figure 52 are plots of the probability that a listener will hear the vowel /a/ or /æ/ at each step in the continuum. In the functions shown in the right panel of Figure 51, for example, /a/, as in “hot” or “sock,” was heard approximately 100% of the time at steps 1 and 2, and then a progressively declining percentage of the time thereafter. The percentage of /æ/ judgments is complementary, starting at zero, and then incrementing gradually to 100% by step 6-7.

Psychometric functions were plotted for each group of respondents by each stimulus type. Figure 51 shows a side-by-side view of the mean responses given by the LM subjects depending on whether the precursor phrases were spoken by Talker LM or Talker UP. The psychometric functions can be seen to be “well-formed” with the /a/~/æ/ response functions changing monotonically along the stimulus continuum. Well-formed psychometric functions are a prerequisite for further investigation of the results. The cross-over point (marked by the dashed line) between the categories /a/ and /æ/ is different for each type of precursor phrase. It is approximately at step 4 for UP precursors, and about 5.5 for LM precursors. Hence, this group shifted their category boundary perceptions by about 1.5 steps along the  $F_2$  continuum in response to the change of precursor talkers.

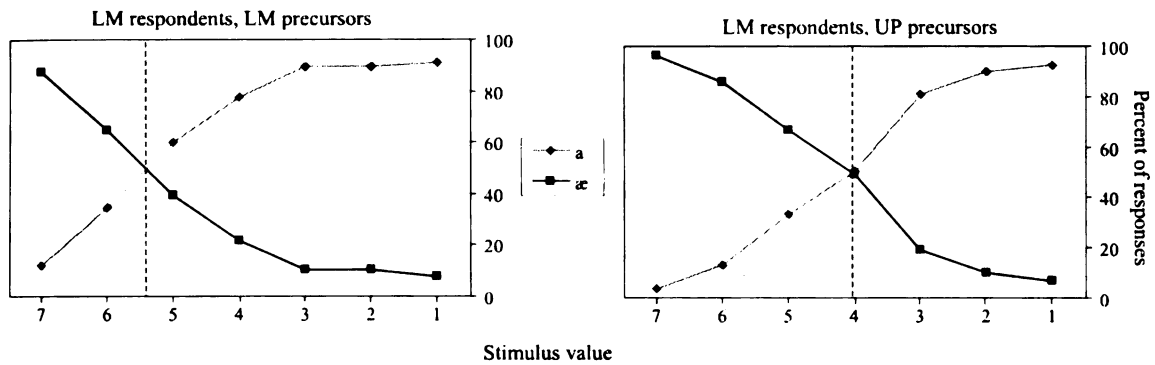


Figure 51. Psychometric functions of LM respondents to the stimuli with LM precursors and UP precursors

In contrast, the UP respondents, whose mean responses are given in Figure 52, showed only a minimal shift based on the change of precursor talkers. Their category cross-over points were at approximately steps 4.75 and 4.5 for the UP and LM precursors, respectively.

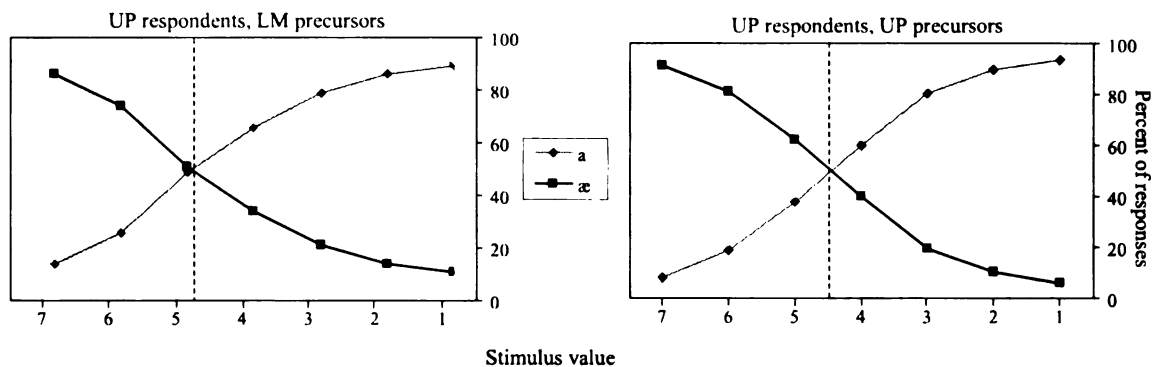


Figure 52. Psychometric functions of UP respondents to the stimuli with LM precursors and UP precursors

### 3.2.7.2 Cross-over points

The notion of the psychometric “cross-over point” is crucial to the understanding of this study. The plots in Figure 51 and Figure 52 show the stimulus values increasing incrementally right-to-left, unlike most mathematical plots. The x-axis was reversed in

order to illustrate the “fronting” *F2* movement of /a/ towards the category /æ/. The synthetic target words differed from one another by 33 Hz along each step of the *F2* continuum. At each trial, the subjects chose either the category /a/, if they thought they heard the word “sock” or “hot” or the category /æ/, if they thought they heard the word “sack” or “hat.” The /a/~ /æ/ category cross-over points are, therefore, a measurable and quantifiable index of the respondents’ “shifting” perceptions across the two categories (Table 8).

Stimulus type	Listener's region	Mean stimulus value	Std. deviation
hot/hat + LM precursor	LM	4.900	1.2135
	UP	4.113	1.8857
hot/hat + UP precursor	LM	3.756	1.0596
	UP	3.988	1.4597
sock /sack+ LM precursor	LM	5.850	.5579
	UP	4.787	.7107
sock /sack+ UP precursor	LM	4.844	.8263
	UP	4.489	.5079

Table 8. Mean stimulus values by stimulus type and listener region

### 3.2.7.3 Choosing a statistical analysis

The cross-over points were identified for each subject in each experimental condition. These values were then analyzed statistically with the General Linear Model (GLM), Repeated Measures (RM) ANOVA procedure (SPSS) to test the hypothesis that cross-

over points varied across the two groups of respondents depending on the dialect of the precursor phrase.

#### **3.2.7.4 The factors**

Factors in the ANOVA were “respondent group” (LM respondents or UP respondents), “precursor talker” (Talker LM or Talker UP), and “word” (“hVt” or “sVk”).

#### **3.2.7.5 Results of the ANOVA test**

Significant main effects were found for word  $F(1,16)=9.543$ ,  $p<.025$  and for precursor type  $F(1,16)=15.079$ ,  $p<.001$ . There was also a significant interaction and for the respondent group and precursor type,  $F(1,16)=6.789$ ,  $p<.05$ .

#### **3.2.7.6 Follow-up paired-samples *t* tests**

Bonferroni-corrected paired-sample *t*-tests were performed to test the nature of the relationship between the respondents’ region and their responses under the different experimental conditions. It was found that LM respondents gave significantly different vowel identity judgments as a function of the precursor phrase,  $t(8)=4.344$ ,  $p<.025$ , unlike the UP respondents, whose responses were not significantly different,  $t(8)=.961$ ,  $p=.365$ . The dashed line in Figure 53 indicates the overall responses of the UP group. The mean stimulus value (category cross-over point) is slightly smaller for UP precursor than for LM precursors. However, this difference is not statistically significant, and the line appears almost horizontal. The LM group (solid line) gave very similar vowel identity judgments to the target words preceded by UP precursors, but they significantly shifted these judgments when the synthetic target word was preceded by a precursor spoken by a



member of their own speech community. The solid line in Figure 53 runs at an almost 45-degree angle, indicating a significant change in the perceptions of the /a/~/æ/ category boundary.

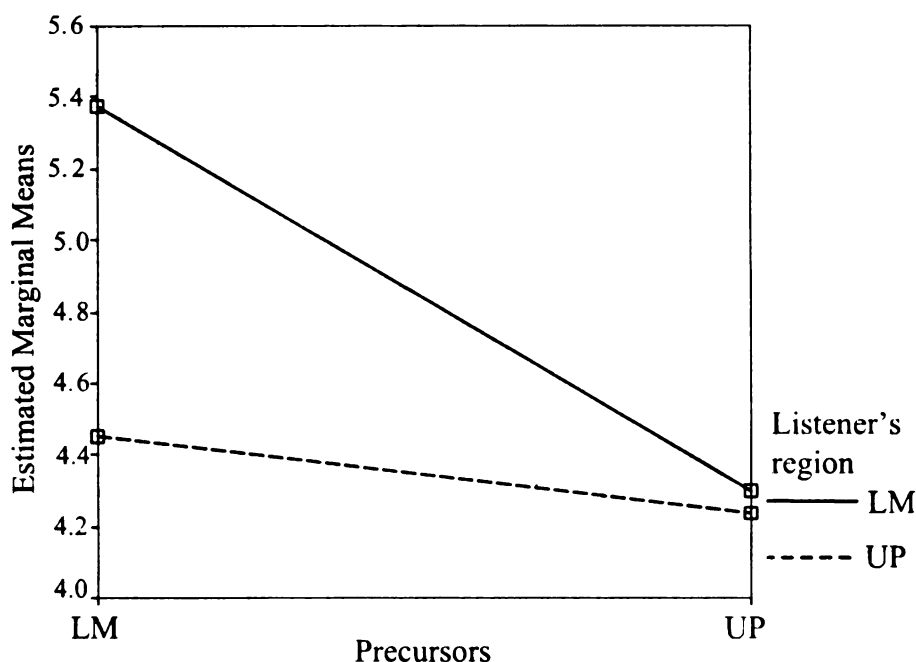


Figure 53. LM and UP responses as a function of the precursor phrase

### 3.2.7.7 The word effect

Next, the word effect was investigated. Overall, the pairs hot/hat~sock/sack triggered significantly different responses,  $t(17)=-3.118, p<.025$ . Figure 54 shows a graph of mean cross-over points by respondent region and by word. Responses to the sequence “hVt” (left panel) were different from those to the “sVk” continuum (right panel). However, despite the differences, these two sequences show very similar relative patterns across the two groups of respondents. The significant word effect remains unexplained. Attempts to correlate it with real-world  $F1$  and  $F2$  of “hat” and “sack” did not yield definitive results.

Future study of this effect could be done by studying the effects of initial consonant duration and formant transitions into the following vowel.

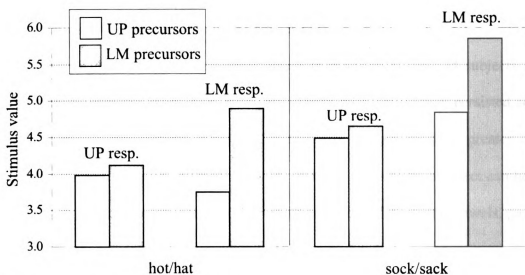


Figure 54. Summary of the word effect

### 3.2.7.8 Precursors

There were four different precursor phrases used in the study (see 3.2.5.3). It was of interest to see whether the results were similar for all four precursor phrases. The precursors varied in overall length and, more importantly, in the phonetic environment immediately preceding the synthetic target word, which could potentially influence the degree to which dialectal information was conveyed to the listener. It is an open question as to how dialectal content was conveyed by the precursors, and how much change in auditory input was necessary to “trigger” a significant perceptual response (see, for example, 1.4.2.2). Those are, after all, some of the most interesting questions in sociophonetics. All four precursor phrases were found to trigger very similar perceptual effects. Statistical analysis revealed no significant difference among them. The circle in Figure 55 represents precursors spoken by Talker UP and the arrow represents precursors

spoken by Talker LM. P1 through P4 indicate the four different precursor phrases. The numbers 3 through 6 along the bottom of the illustration denote the middle portion of the 7-step continuum. The length of the arrow shaft indicates the difference in response to the target words preceded by UP and LM carrier phrases. In the case of UP listeners (right), the arrow shafts are consistently short, while in the case of the LM subjects, they are consistently long, indicating that, while the overall perceptual effect persisted across the two groups of respondents, no specific precursor phrase made a greater or lesser contribution to the results. These results suggest that the perceptual effect can be elicited fully by any precursor that offers a reasonable sampling of a talker's vowels, particularly /æ/, /a/, and /ε/.

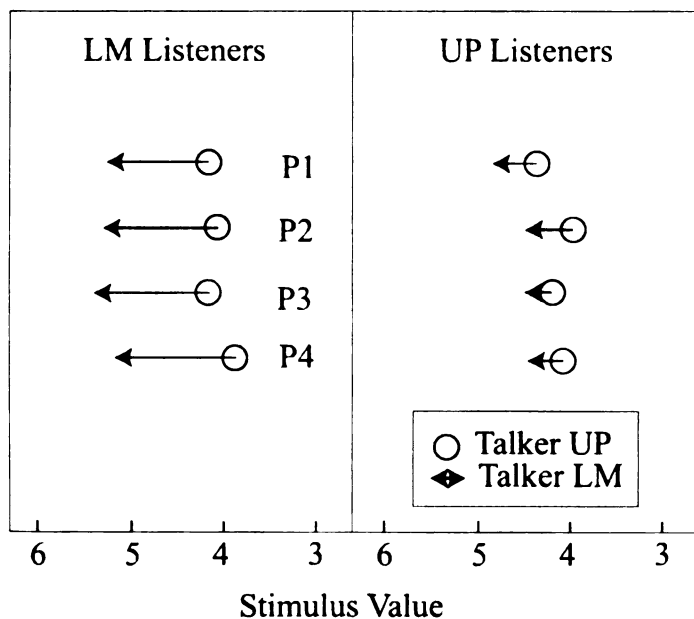


Figure 55. Precursor phrases 1 through 4 (P1-P4) - not significant in vowel identity judgments (From Rakerd and Plichta (2003) with permission)

### **3.2.8 Summary**

The research questions of 3.2.1 (repeated here) were evaluated with empirical research and quantitative analysis.

1. Does talker-dependent, sociolinguistic information influence speech perception?
2. Do hearer-dependent, sociolinguistic factors influence speech perception?
3. What is the nature of talker normalization involved in this process?
4. Is /a/-fronting perceptually salient among NCCS speakers?

The short answer to questions 1, 2, and 4 is “yes.” Answering question 3 is more complex. Below, is a summary of the major findings of the /a/-fronting study.

#### **3.2.8.1 Summary point #1**

Dialectal information contained in the precursor phrase plays a significant role in vowel identity judgments. Undoubtedly, LM respondents were influenced by the dialect of the precursor phrase. They shifted their perception of the /a/~ /æ/ boundary toward higher *F*<sub>2</sub> frequencies, or toward the front the two-dimensional vowel space.

#### **3.2.8.2 Summary point # 3**

As evidenced in 3.2.7.8, specific precursor phrases overall were not significant.

#### **3.2.8.3 Summary point # 4**

The pairs hot~hat and sock~sack triggered significantly different responses. However, these responses were relative to each group of respondents' overall vowel identity judgments.

#### **3.2.8.4 Summary point # 5**

In single-word conditions, respondent region was not significant. This seems to indicate that the LM and UP speech communities shared a general contrastive inventory of /a/ and /æ/, but that the range of phonetic (or allophonic) representations available to the Detroit community was significantly broader than that of the UP community (see 4.3.2 for a discussion on the role of multiple allophonic representations in a sociophonetic model of sound change).

## **CHAPTER 4 ON HEARER-MEDIATED SOUND CHANGE**

### **4.1 HOW DO DIALECTS DIFFER?**

One of the fundamental questions in sociolinguistics is how dialects differ from one another. Inevitably, this question leads to the more specific problem of sound change. Why is it that certain speech communities develop certain pronunciation patterns? In the field of sociolinguistics, this question has been approached primarily from the point of view of speech production, with an emphasis on the role of the speaker.

#### **4.1.1 Speaker-mediated sound change in dialectology**

Most typically, sociolinguists assume that the inherent linguistic variability and social aspects of language use (particularly non-standard usage), are the main forces behind sound change. Some sociolinguists believe that sound change occurs as a result of language use economy or ease of production (Kroch, 1978), while others argue that it evolves directly from the talkers' linguistic competence. Chambers (1995) argues:

“Where change is involved, a certain variant will occur in the speech of children, though it is absent in the speech of their parents, or, more typically, a variant in the parents' speech will occur in the speech of their children with greater frequency, and in the speech of their grandchildren with even greater frequency. .... The logical conclusion, as time goes by, will be the categorical use of that new variant and the elimination of older variants” (p. 185).

#### **4.1.2 The need for a broader model of sound change**

While one has to agree with Chambers' scenario, one must also note that this model misses one, crucially important, element – the hearer. For any viable sound change

model, there need to be at least three elements (1) the speaker, (2) the hearer, and (3) language variation in the speech community. Therefore, a satisfactory research paradigm would have to be set at the interface of speech production (the speaker's role), speech perception (the hearer's role), and sociophonetics (language variation in the acoustic/phonetic domain).

#### **4.1.2.1 Speech production and linguistic variability**

It is no secret that sociophoneticians think of the articulatory-acoustic properties of language use very seriously. While this may have little appeal to formal linguists, it certainly is an interest shared with speech scientists.

Speech scientists think of speech production as a complex system of rapid articulatory gestures, subject to physical limitations of human physiology (e.g., Faber (1992)). This physiological aspect of speech production is crucial to the understanding of the role of the speaker in sound change, as human physiology is believed to be the most important source of variability (see, for example, the discussion on VTL variability in 3.1.1.1).

Sociolinguists take a different approach. They do believe in the abstractness of language and linguistic representation. For example, vowel shifts are described as movements within the two-dimensional space delimited by objective formant values, but classified by theory-dependent binary representations, such as "tense" versus "lax" (Labov, 1991). Sociolinguists also think of variability as a constituent of sound change, but instead of focusing on physiology, they are more interested in socially constructed patterns of

linguistic variability. Some also believe that this type of variability can be predicted from a general linguistic theory (e.g., Labov (1994)).

#### **4.1.2.2 Speech perception and linguistic variability**

Most speech perception theories agree that the relationship between the acoustic stimulus and a related auditory response is not linear (see 1.4.2.2). They also believe that speech perception is, to a large degree, categorical (see 1.4.4.3). Sociolinguists take a different approach. Perceptual dialectologists study beliefs and attitudes while sociophoneticians are most interested in “within category” perception (see 1.4.2.3).

#### **4.1.2.3 Why current models cannot account for nasalization and sociophonetic perceptual category boundary shifting**

##### ***4.1.2.3.1 Problems with nasalization***

As evidenced in CHAPTER 2, coarticulatory vowel nasalization occurs as a result of specific gestures of the human speech production apparatus. It can be measured, either from spectra, or aerodynamically, and its acoustic properties can be predicted from a general theory of speech production. Coarticulatory vowel nasalization in Michigan has a physiological reality (but is non-contrastive and occurs in non-pathological speech and in non-nasal environments) and a sociophonetic reality (but it has no place in existing sociolinguistic taxonomy). The problem becomes even more complex in light of the fact that nasalization can be predicted from talker participation in NCCS.

##### ***4.1.2.3.2 Problems with sociophonetic perceptual category boundary shifting***

The perceptual category shifting depending on dialectal input (CHAPTER 3) is also problematic for current speech and language theories. As you recall, only Lower



Michigan respondents shifted the /a~/æ/ boundary as a function of dialectal input. Speech perception theories cannot fully account for talker normalization mediated through the hearer's speech community.

Sociolinguistic theory has trouble accounting for this phenomenon as well. This particular type of perception cannot be, for example, explained by folk linguistics (in ways that Niedzielski's study cited in 1.4.2.1 could). One cannot attribute the results to the perceivers' beliefs about standardness or correctness.

## **4.2 NASALIZATION, /a/-FRONTING, AND SOUND CHANGE**

### **4.2.1 Articulatory variability and sound change**

Vowel nasalization and perceptions of /a/-fronting in Michigan are correlated with NCCS. These two studies provide evidence that sound change has a significant hearer-mediated component. Even though this idea is not new to speech science, there has not been an account that would consider variability in broader, sociophonetic terms. For example, Ohala (1981, 1993, 1996), suggested that sound change occurs when listeners misapply corrective rules that serve to correct phonetic variability. He defines variability as "noise," or variability primarily due to articulatory factors.

Faber (1992) is distrustful of Labovian sociolinguistics and argues that before phonetic variants can spread, they must first be salient and reproducible. Faber also seems to think of acoustic variability only in a purely articulatory/phonetic sense and attributes sound change mostly to the relationship between articulatory variability and categorical speech perception.

### **4.2.2 Vowel nasalization and perceived height**

The first formant frequency in Hertz, as relative to tongue height, is often referred to as “vowel height.” Several studies in speech perception (e.g., Beddor, Krakow, & Goldstein, (1986)) have demonstrated that perceived vowel height does not always correspond to the actual value of  $F1$ . This is particularly true of nasalized vowels.

Krakow et al (1988) demonstrated that listeners can be confused by the similarity of nasalization and tongue height effects. Krakow and her colleagues used synthetic speech samples with varying degrees of velopharyngeal opening to simulate the effects of changes in tongue height. They also discovered that extra coarticulatory effects from the preceding and following environment influenced the perceptions of vowel height.

Beddor and Hawkins (1990), in a series of perceptual experiments with synthetic speech, discovered that vowel height is determined by the most prominent harmonics in the low-frequency region of the spectrum, as well as by the slopes of the skirts of those harmonics. This finding implies that the relationship between nasalization and perceived vowel height is continuous. The stronger the nasalization, the higher the amplitude of  $F1_n$ , the more significant a role it will play in perception. Similarly, when formant bandwidths are wide, perception will be determined more by the shape of the spectrum than by any specific low-frequency harmonics.

Beddor and Hawkins discovered that listeners would always average their perception between the first oral and nasal formants. This, of course, means that low vowels (such as /æ/) would be perceived higher, and high vowels such as /ε/, would be perceived lower.

Beddor and Hawkins have also found the critical distance of 3.5 Bark to play a role in the perceptions of vowel height. Kingston (1991) provided additional support for the claim that perceived vowel height is a function of several covarying articulations. He presented quantitative evidence for the fact that nasalization and  $F_0$  are perceptually integrated with the acoustic effects of tongue height.

The findings obtained in CHAPTER 2 and the perceptual experiments discussed above can be summarized as follows:

1. Perceived vowel height is not a sole function of  $F_1$ . It is influenced by a bundle of covarying articulatory gestures, out of which the opening of the velopharyngeal port is the most prominent.
2. Vowel nasalization in Michigan is distributed along NCCS lines of region and sex.
3. In NCCS, /æ/ is considered raised, and /ε/ is considered lowered.
4. The presence of a nasal peak may cause the NCCS-influenced /æ/ to be perceived higher and /ε/ to be perceived lower - subject to auditory constraints of the critical distance.
5. The simultaneous raising of /æ/ and lowering of /ε/ is hearer-mediated in NCCS.

In the course of this chapter, evidence has been provided to support the claim that certain NCCS vowel movements, such as the raising of /æ/ and the lowering of /ε/, are mediated through the hearer's perceptions of vowel height. Figure 56 shows a Principle Components Analysis vowel plot of normalized NCCS vowels collected from 26 Lower Michigan talkers. Note that the vowels /ε/, /æ/, and /ʌ/ share a fair amount of discriminant space. The  $F_1/F_2$  classificatory system has failed, but one still must be able to account for the perceived contrast between these three vowels (also see Figure 42).

Clearly, there must be other articulatory-acoustic features that make it possible for hearers to unambiguously identify these vowels.

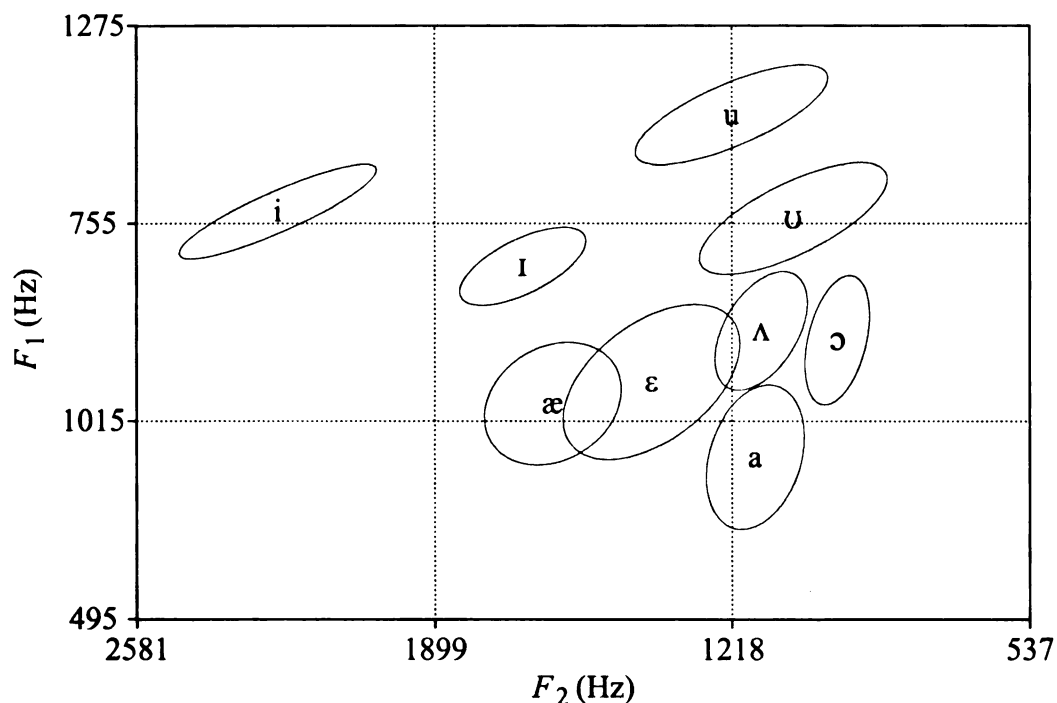


Figure 56. Normalized Principal Components plot of an NCCS population from Lower Michigan

At this stage, one can formulate the claim that nasalization is actively re-shaping the NCCS vowel space. Nasalization and tongue position are both integrated in perceived vowel height. The  $F_1/F_2$  taxonomy (Figure 56) is not able to capture this coarticulatory process. Speakers reproduce their own auditory perceptions of the vowels /æ/ and /ε/ by simultaneously opening the velopharyngeal port and changing tongue height. It might be possible that, currently, speakers and hearers are using both strategies to negotiate the desired perceptual effect.

#### **4.2.2.1 Will nasalization prevail?**

It is not unreasonable to argue that the NCCS vowel system will not end up consisting exclusively of nasal vowels. It is no accident that nasal vowels are much less common in world languages than oral vowels. Wright (1986), for example, demonstrated that, in experimental conditions, all-nasal systems suffered a significant reduction in vowel contrast. Still, nasalization can be expected to continue playing an important role in altering the vowel space in NCCS-influenced dialects of English.

Which occurs first in NCCS, nasalization or oral cavity gestures? While one might never know for certain, one can speculate that it is likely that nasalization occurs first, as a coarticulatory process. It probably occurs as a function of both physiological (low vowels are more likely to be nasalized) and sociophonetic (as a sociolinguistic marker) reasons. Because of its contrast reducing properties, it may be forcing speech communities to renegotiate changes in oral articulations (such as tongue height). As a result, a system-wide vowel shift, such as NCCS, may finally take place.

#### **4.2.3 Hearer-mediated /a/-fronting**

##### **4.2.3.1 Information integration in speech perception**

Integrating top-down and bottom-up cues in speech perception is not a new idea (see, for instance 3.1.1.3). However, the /a/-fronting study has broadened this concept by adding to it a strong dialectal component and demonstrating that perceivers rely on integrating input-dependent sociophonetic content, as well as, mediating their perceptions via the sociophonetic experience of their speech community.

#### **4.2.3.2 Perceptual salience**

The concept of perceptual salience is fundamental to the theory of sound change. When individuals with different dialects live in close contact, they may adopt some of the features of that other dialect (e.g., Kerswill (1994)). A sociolinguistic theory of dialect contact or dialect accommodation must be able to account for the fact that only specific features are perpetuated, while others die out. Most typically, such accounts would be based on a combination of synchronic and diachronic studies of production data. Some of the sound change phenomena would be determined by language universals, while others would be considered unpredictable or idiosyncratic.

The /a/-fronting study provides a quantificational account of perception data. It allows one to determine the micro-level acoustic/phonetic detail that carries dialectally salient weight. In addition, we discover that sound change is not driven exclusively by abstract language universals, but it is determined by the capabilities of the human auditory system and the hearer's own dialect history.

#### **4.2.3.3 33 Hertz and abstract features**

Labov (1991) discusses a theory of on-going vowel shifts. One of his most important claims is that certain properties of vowel shifts can be predicted from a set of generalized principles, such as “low nonperipheral vowels become peripheral.” For Labov, features such as, +/- peripheral have their measurable acoustic correlates obtained from acoustic measurements of speech production. For instance, the forward periphery, he argues, follows the line of  $F2-2F1=C$ , where C is a constant that varies with the speaker.

Even if one accepts Labov's definition of vowel shifts, one still finds it inadequate to account for the results of the /a/-fronting study. The study demonstrated that vowel movements are crucially hearer-mediated and cannot be predicted exclusively from production data. It is the hearer and their speech community who negotiate the acoustic properties of vowel shifts. In this particular case, the Lower Michiganders were sensitive to changes in *F*<sub>2</sub> of as little as 33 Hz, while the UP respondents would have required a significantly larger interval before they could process this input as meaningful. The sociophonetically relative perceptual nature of absolute formant frequencies demonstrated by the study provides yet another argument in favor of augmenting existing theories of sound change by the hearer-negotiated component.

### **4.3 TOWARD A SOCIOPHONETIC MODEL OF HEARER-MEDIATED SOUND CHANGE**

#### **4.3.1 Phonological models of nasalization**

As mentioned in 4.2, Ohala (1981, 1993, 1996) put forth a model of sound change based on the negotiation of ambiguity and the interpretation of articulatory "noise" between the speaker and the hearer. For example, vowels followed by nasal consonants, /vn/, can be progressively interpreted as /*ṽ*/ if the listener discontinues the application of a "corrective rule," necessary for the interpretation of the nasal environment in /vn/. The Ohalan model, as applied to nasalized vowels, can be summarized in Figure 57 below. When the vowel /a/ is pronounced in a nasal environment, it is subject to vowel nasalization (the lowering of the velum), which the listener initially "corrects" for and identifies the vowel as /a/ and the following consonant as /n/. However, once the listener stops applying the corrective rule, the vowel is perceived as /*ṽ*/ followed by silence. Thus perceived vowel

nasalization becomes more salient and the listener begins reproducing and spreading the new variant /ã/. Even though this particular type of /n/-deletion has significant phonological consequences (the emergence of the new contrastive category /ã/), Ohala does not believe that there is an intermediate phonological derivation between abstract representation and acoustic-phonetic output.

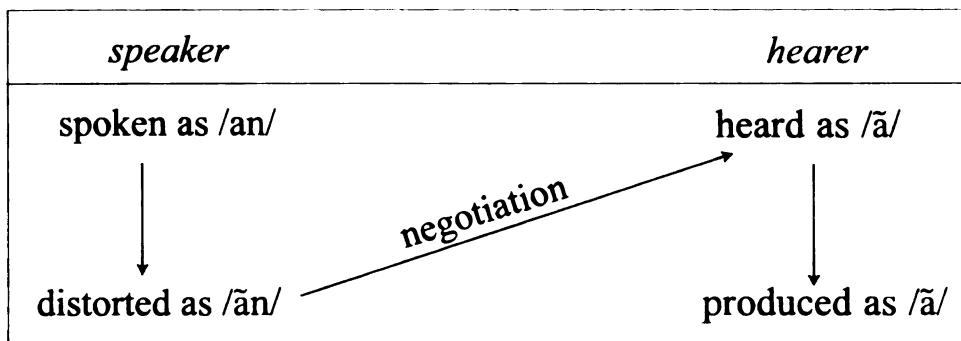


Figure 57. Summary of the Ohalan listener-mediated model of sound change

Hajek (1997) modified the Ohalan model of sound change by adding to it a phonological derivational component related to the grammatical structure predicted from the theory of Lexical Phonology. At the first stage of sound change, listeners deal with the ambiguity of the speech signal by reinterpreting or misinterpreting the nasal environment directly following the vowel (/vn/). This occurs as a result of language-specific rules and can lead to further contextual reinterpretation of vowel nasalization until the process becomes phonetically stable and interpretable at a deeper level of the post-lexical component. As a result, a phonemically nasal vowel, /ã/, may emerge both at deep and surface levels of grammar.



### 4.3.2 A sociophonetic model

The studies in CHAPTER 2 and CHAPTER 3 provided evidence that there are significant sources of variability in the speech signal related to sociolinguistic factors such as region and sex. Both vowel nasalization and the perceptual /a~/æ/ category shift can also be predicted from talker participation in NCCS. The two studies showed that the speech signal is rich in sociolinguistic information and that this information is actively used in speech production and perception. Figure 58 shows a proposed model of sociophonetic sound change whereby articulatory gestures and their perceptions are mediated via the speaker, the hearer, and their speech community. For example, as evidenced in CHAPTER 2, certain Lower Michigan individuals are likely to pronounce the vowel /æ/ in the word “bag” with the velopharyngeal port open, which is not caused by the presence of a nasal consonant, as seen in Ohala’s model in Figure 57 (articulatory variability), but, rather, by sociolinguistic forces specific to their speech community (sociophonetic variability). At the same time, such pronunciations are subject to negotiation and reinterpretation by the hearer and their speech community as well. As a result, the output of such reinterpretation may vary – for some hearers it will result in a nasal vowel (e.g., /bæ̃g/), while for others, the vowel will be pronounced with a lowered *F*<sub>1</sub> (e.g., /bɪæ̃g/). Such vocalic instability is a characteristic feature of early stages of vowel shifts (Labov et al., 1997; Stockwell & Minkova, 2000).

Similarly, Lower Michigan talkers are likely to pronounce the word “block” with a fronted *F*<sub>2</sub> (Figure 58), which makes this word sound more like “black” ([blæk]) to non-NCCS speakers (such as the UP respondents described in 3.2.4). Lower Michigan hearers

are able to normalize the increased frequency of *F2* (see 3.2.8) and correctly identify the “shifted” vowel as /a/. At some point in this speaker-hearer negotiation, the fronted /a/ becomes perceptually salient in their speech community (see, for example, the results of the a-fronting study in 3.2.8), and the previously non-standard, “fronted” pronunciation [blæk] becomes the norm.

In of the studies presented CHAPTER 2 and CHAPTER 3, the respondents’ speech community acts as a dialectal filter in the speaker-hearer negotiation. This filter “allows” or “blocks” the use (speech production) and comprehension (speech perception) of non-standard forms by the speaker and the hearer.

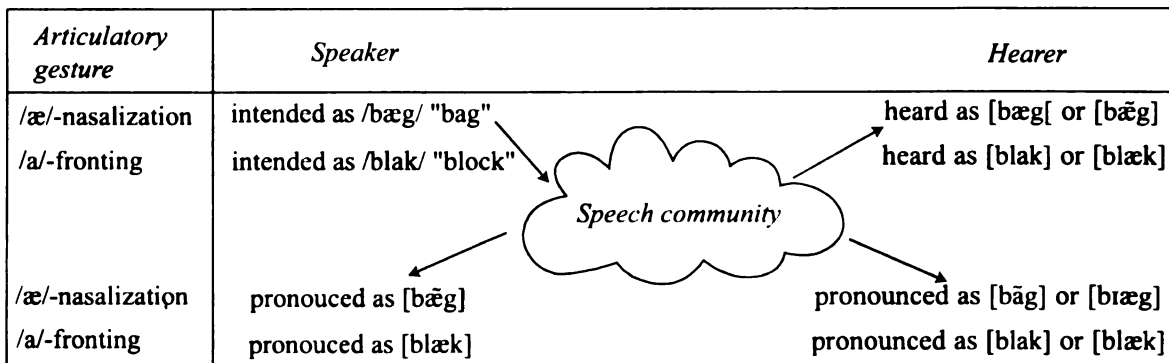


Figure 58. Sociophonetic model of hearer-mediated sound change

It is, therefore, not unreasonable to argue that a successful model of sound change must be able to account for sociophonetic variability (see 4.1.2) regardless of whether it assumes an intermediate phonological derivational component or not. NCCS speech communities are reinterpreting and misinterpreting, negotiating and renegotiating vowel nasalization, tongue height, tongue frontness, and, possibly, a host of other articulatory-phonetic features of vowels. As a result, the diverse and dynamic speech communities of the American North, such as Detroit-area European Americans, are quickly adopting the

new vocalic changes. NCCS is, therefore, likely to continue playing an important role in modern American English.

## APPENDIX A - WORDLIST USED IN CHAPTER 2

1. jaw	26. back
2. job	27. bit
3. knock	28. book
4. lid	29. bond
5. lot	30. buddy
6. nasty	31. but
7. pot	32. sad
8. set	33. cap
9. shed	34. cot
10. bob	35. dad
11. shot	36. took
12. sit	37. left
13. kick	38. bet
14. nag	39. dead
15. man	40. did
16. caught	41. hat
17. head	42. sat
18. cod	43. hit
19. coat	44. should
20. sought	45. shut
21. test	46. hook
22. hut	47. hot
23. boat	48. cat
24. but	49. mat
25. bag	50. cut

Table 9. Wordlist used in studies described in CHAPTER 2

APPENDIX B – QUESTIONNAIRE (CHAPTER 2 and CHAPTER 3CHAPTER 2)

What is your age?

---

What is your current occupation?

---

What are your parents' occupations?

---

What is your education?

---

Where do you live?

---

Tell me about your neighborhood

---

The most common type of housing here is detached house (164/558). Is this you?

---

Have you moved a lot in your life? Where? Other cities, states, countries?

---

Are there any family members living in this area?

---

Are there many people from your neighborhood in your workplace?

---

Are there more men or women in your workplace?

---

Are there any activities in the neighborhood that you participate in? (sports, park, community center, gym, etc.)

---

Do you hang out with your work mates after work?

---

Do you like living in your city/town/village?

---

If you have a choice, would you live some place else?

---

What would be one thing you would like to change about this town?

---

# APPENDIX C – SUMMARY OF INDIVIDUAL CASES OF %N

Region	Sex	Subject Code	Mean %N	N	Std. Deviation
LM	female	LM30	22.948	48	8.901398704
		LM31	23.569	47	6.648855486
		LM39	38.326	48	8.931787802
		LM40	18.994	47	6.337375243
		LM49	25.214	48	12.20337446
		LM50	26.408	48	11.682937
		LM55	21.703	46	10.79227997
		Total	25.355	332	11.14084192
	male	LM32	18.625	48	8.701631786
		LM34	11.852	47	8.159063948
		LM35	14.268	48	7.671490925
		LM36	25.582	45	5.943389486
		LM51	13.332	26	6.061097387
		Total	16.732	214	7.307334707
		MM	female	LM37	12.722
LM43	15.174			47	8.905235309
LM44	20.142			45	11.71140184
LM46	18.876			48	6.292540614
Total	16.725			186	9.232970837
male	LM28		9.1647	45	5.280797374
	LM38		8.9396	48	4.830615554
	LM41		8.1839	46	5.278136236
	LM60		8.5272	47	4.769683565
	Total		8.703	186	5.013505799
UP	female	LM33	12.219	46	7.144046909
		LM42	10.753	46	6.503185693
		LM45	17.558	46	8.68847108
		LM47	12.505	43	7.587818668
		LM48	15.78	47	7.294982194
		LM53	10.683	47	6.991691064
		Total	13.258	275	7.768383017
	male	LM56	8.6149	47	5.647302885
		LM29	10.08	47	6.824183964
		LM58	9.3065	48	5.109943785
		LM61	9.3156	48	5.937313278
Total	9.3294	190	5.879685978		

Table 10. Summary of individual cases for %N

## APPENDIX D – CONSENT FORM CHAPTER 2

Vowel nasalization among Michigan speakers of English  
Bartek Plichta, Department of Linguistics and Languages,  
Michigan State University  
East Lansing, MI 48824  
plichtab@msu.edu  
tel. 517-3559300

### Consent Form

This study is done as part of the research project in the area of speech production in the context of on-going dialectal changes in the state of Michigan. The purpose of this study is to investigate how Michiganders produce several different sounds of American English.

The subjects will be asked to put on a small, head-worn microphone and read a list of 50 simple, monosyllabic words. The investigator will tape record those words. In some cases, the subjects may be asked to speak to a non-invasive device called a “nasometer,” which is a microphone with a small, plastic mask around it.

The subjects will be asked to provide basic demographic information, including their age, gender, ethnicity, and socio-economic status.

Participation in the study is voluntary. The subjects may refuse to participate at any stage of the study.

The subjects' privacy will be protected to the maximum extent allowable by the law.

There is no risk of physical injury involved in the study.

The subjects may contact:

Professor Dennis Preston  
at the Department of Linguistics and Languages (Michigan State University)  
A-740 Wells Hall, tel. 517 353-9945; email: preston@msu.edu

or Ashir Kumar, M.D., Chair of the University Committee on Research Involving Human Subjects (UCRIHS)

phone: (517) 355-2180, fax: (517) 432-4503, e-mail: ucrihs@msu.edu,  
regular mail: 202 Olds Hall, East Lansing, MI 48824.

I have read, understood, and accepted the consent form:

Date

Signed

## APPENDIX E – CONSENT FORM CHAPTER 3

### *Perceptions of /a/-fronting across two Michigan dialects of English*

#### Consent Form

This survey investigates perceptions of the fronting of the vowel /a/ (in words, such as *hot, sock, top, lot*, etc.)

You will be asked to sit in front of a multimedia computer.

You will read and accept (or not) the consent form.

You will be given brief instructions on how to proceed.

You will be asked to put on headphones.

The experiment takes about 45 minutes to complete.

The experiment will consist of 2 blocks with a break in between. You will be automatically notified when the break should occur.

You can interrupt the experiment at any point, should you require an extra break.

In each block, you will hear a short sentence, at the end of which, you will hear the word "sock" or "sack." Your task is to click the left button if you hear "sock" and the right mouse button if you hear "sack."

Each block consists of 224 trials.

Participation in the study is voluntary. You may refuse to participate at any stage of the study.

Your privacy will be protected to the maximum extent allowable by the law.

There is no risk of physical injury involved in the study.

If you have questions about the study, contact:

Brad Rakerd, Professor, Dept. of Audiology & Speech Sciences

373 Communication Arts Building

Michigan State University

East Lansing, MI 48824-1212

phone: (517) 432-8195; fax: 432-2370

In case you have questions or concerns about your rights as a research participant, please feel free to contact Peter Vasilenko, Ph.D., Chair of the University Committee on

Research Involving Human Subjects (UCRIHS)

phone: (517) 355-2180, fax: (517) 432-4503, e-mail: [ucrihs@msu.edu](mailto:ucrihs@msu.edu),

regular mail: 202 Olds Hall, East Lansing, MI 48824.

By signing this form, I volunteer to participate in this study.

YOUR NAME (please print)

SIGNATURE



# APPENDIX F – WORDLIST USED IN CHAPTER 3

---

1. jaw	26. move
2. job	27. bit
3. knock	28. book
4. lid	29. boot
5. lot	30. rude
6. nasty	31. but
7. pot	32. sad
8. set	33. cap
9. shed	34. cot
10. heat	35. dad
11. shot	36. bead
12. sit	37. left
13. soothe	38. bet
14. nag	39. dead
15. man	40. did
16. caught	41. hat
17. head	42. sat
18. cod	43. hit
19. coat	44. should
20. sought	45. shut
21. test	46. hook
22. hut	47. hot
23. wheat	48. cat
24. but	49. mat
25. bag	50. cut

---

Table 11. Wordlist used in the study in CHAPTER 3

## BIBLIOGRAPHY

- Abramson, A. S., & Lisker, L. (1968). Voice timing: Cross-language experiments in identification and discrimination. *Journal of the Acoustical Society of America*, 44(1), 377.
- Adank, P. (2003). *Vowel Normalization: a perceptual-acoustic study of Dutch vowels*. Unpublished Ph.D. dissertation, University of Nijmegen, Nijmegen.
- Anderson, B. L. (2002). Dialect leveling and /ai/ monophthongization among African American Detroiters. *Journal of Sociolinguistics*, 6(1), 86-98.
- Bailey, G., & Thomas, E. (1998). Some Aspects of African-American Vernacular English Phonology. In S. Mufwene, J. Rickford, J. Baugh & G. Bailey (Eds.), *African American English* (pp. 85-109). London: Routledge.
- Barlett, B., & Barlett, J. (1998). *Practical recording techniques* (Second ed.). Boston: Focal Press.
- Beckman, M., & Hirschberg, J. (1994). *The ToBE Annotation Conventions*. Unpublished manuscript, Ohio State University.
- Beddor, P. S., & Hawkins, S. (1990). The influence of spectral prominence on perceived vowel quality. *Journal of the Acoustical Society of America*, 87(6), 2684-2704.
- Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook*, 3, 197-217.
- Boersma, P., & Weenink, D. (2002). Praat (Version 4.2).
- Chambers, J. (1995). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. Basil: Blackwell.
- Chen, M. (1995). Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *Journal of the Acoustical Society of America*, 98(5, Pt 1), 2443-2453.

- Chen, M. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America*, 102(4), 2360-2370.
- Chistovich, L. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 77, 789-805.
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- Eckert, P. (1999). *Linguistic Variation as Social Practice*. Oxford: Blackwell.
- Ee Ling, L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable timing in Singapore English. *Language and Speech*, 43(3), 377-401.
- Evans, B. (2001). *Dialect accommodation and the Northern Cities Shift*. Unpublished Ph.D. dissertation, Michigan State University, East Lansing.
- Evans, B., & Preston, D. (2000). *When being normal isn't nice*. Paper presented at the New Ways of Analyzing Variation 29 conference, Michigan State University, East Lansing.
- Faber, A. (1992). Articulatory Variability, Categorical Perception, and the Inevitability of Sound Change. In G. W. Davis & G. K. Iverson (Eds.), *Explanation in Historical Linguistics* (pp. 59-75). Amsterdam: Johns Benjamins.
- Fitch, H. L., Hawles, T., Erickson, D. M., & Lieberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop consonants manner. *Perception and Psychophysics*, 27, 343-350.
- Fridland, V. (1998). *The Southern Vowel Shift: Linguistic and Social Factors*. Unpublished Ph.D. dissertation, Michigan State University, East Lansing.
- Glottal, E. (2002). OroNasal Mask System (Version 1.5): Avaaz Innovations.
- Gordon, M. (1997). *Urban sound change beyond the city limits: the spread of the Northern Cities Shift in Michigan*. Unpublished Ph.D., University of Michigan, Ann Arbor.

- Graff, D., Labov, W., & Harris, W. (1983). Testing listeners' reactions to phonological markers of ethnic identity: a new method for sociolinguistic research. In D. Sankoff (Ed.), *Diversity and diachrony. Current issues in linguistic theory* (Vol. 53, pp. 45-58). Amsterdam: Benjamins.
- Gurijala, A., Deller, J., & Seadle, M. (2002). *Watermarking through parametric modeling*. Paper presented at the Proceedings of the International Conferences of Spoken Language Processing, Denver.
- Guy, G. R. (1980). Variation in the group and in the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1-36). New York: Academic Press.
- Hajek, J. (1997). *Universals of sound change in nasalization*. Oxford, UK: Boston.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*(97), 3099–3111.
- Huber, D. M., & Williams, P. (1998). Professional microphone techniques. Emeryville, CA: Mix Books.
- Hunt, K. (1996, Aug 27). Sony revives MiniDisc in package deal. *Los Angeles Times*, p. 5:1.
- Ito, R. (2000). *Diffusion of urban sound change: a case of the Northern Cities Shift*. Unpublished Ph.D. dissertation, Michigan State University, East Lansing.
- Jacoby, W. G. (1998). *Statistical graphics for visualizing multivariate data*. Thousand Oaks: Sage Publications.
- Johnson, K. (1989). Contrast and normalization in vowel perception. *Journal of Phonetics*(18), 229-254.
- Jones, D. (1964). *An outline of English phonetics*. Cambridge: W Heffer & Sons Ltd.
- KayElemetrics. (1998). Computerized Speech Lab. Lincoln Park: Kay Elemetrics Corp.

- Kay Elemetrics. (2003). Nasometer II. Lincoln Park: Kay Elemetrics Corp.
- Kerswill, P. (1994). *Dialects Converging: Rural speech in urban Norway*. Oxford: Clarendon Press.
- Kingston, J. (1991). Integrating articulations in the perceptions of vowel height. *Phonetica*, 48, 149-179.
- Klatt, D., & Klatt, L. C. (1990). Analysis, synthesis, and the perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- Krakov, R. A., Beddor, P. S., Goldstein, L. M., & Fowler, C. A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America*, 83, 1146-1158.
- Krakov, R. A., & Huffman, M. K. (1993). Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory. In M. K. Huffman & R. A. Krakow (Eds.), *Phonetics and Phonology 5: nasals, nasalization, and the velum*. Sand Diego: Academic Press.
- Kroch, A. (1978). Toward a theory of social dialect variation. *Language in Society*, 7, 17-36.
- Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45, 715-762.
- Labov, W. (1972). The social stratification of (r) in New York City department stores. In W. Labov (Ed.), *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1991). The three dialects of English. In P. Eckert (Ed.), *Quantitative Analyses of Sound Change* (pp. 1-44). New York: Academic Press.
- Labov, W. (1994). *Principles of linguistic change*. Oxford, UK; Cambridge, MA: Blackwell.
- Labov, W. (2001). *Principles of linguistic change: social factors*. Oxford: Blackwell.

- Labov, W., Ash, S., & Boberg, C. (1997). A National Map of Regional Dialects of American English. Telsur Project.
- Labov, W., Yeager, M., & Steiner, R. (1972). *A Quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.
- Ladefoged, P. (1967). *Three areas of experimental phonetics*. London: Oxford University Press.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 1(29), 99-104.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773-1781.
- Lobanov, B. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49, 606-608.
- Massaro, D. (1998). *Perceiving talking faces: from speech perception to a behavioral principle*. Cambridge, Mass: MIT Press.
- Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85, 2114-2134.
- Milroy, L. (1980). *Language and social networks*. Oxford: Blackwell.
- Moll, K. L. (1962). Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research*, 5(1), 30-37.
- Nearey, T. (1977). *Phonetic features system for vowels*. Unpublished PhD dissertation, University of Connecticut.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088-2113.

- Niedzielski, N. (1999). The effects of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 49-62.
- Nordstrom, P., & Lindblom, B. (1975). A normalization procedure for vowel formant data. *Proceedings of the 8th. International Congress of Phonetic Sciences.*, 212.
- Nussbaum, H., & Morin, T. M. (1993). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech Perception, Production, and Linguistic Structure*. Tokyo: OHM Publishing Company.
- Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hnedrick & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society.
- Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13(1-2, Spec Issue (Oct 1993), 155-161.
- Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, 99(3), 1718-1725.
- Paolillo, J. C. (2001). *Variable rule analysis: using logistic regression in linguistic models of variation*. Stanford, Calif.: CSLI: Cambridge.
- Peterson, G., & Barney, H. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*(24), 175-184.
- Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13, 253-260.
- Plichta, B. (2002). Best practices in the acquisition, processing, and analysis of acoustic speech signals. *U. Penn Working Papers in Linguistics*, 8.3.
- Plichta, B. (2004). Akustyk for Praat (Version 1.7.2). East Lansing: Michigan State University.
- Plichta, B., & Preston, D. (2004). The /ay/s Have It: The perception of /ay/ as a North-South stereotype in US English. In T. Kristiansen, Coupland, Nikolas, Garrett,

Peter (eds) (Ed.), *Acta Linguistica Hafniensia* (Vol. Theme Issue on Subjective Processes in Language Variation and Change).

Pohlmann, K. C. (2000). *Principles of Digital Audio* (Fourth ed.). New York: McGraw-Hill.

Preston, D. (1996). Where the worst English is spoken. In E. Shneider (Ed.), *Focus on the USA* (pp. 297-360). Amsterdam: John Benjamins.

Preston, D. (1999). Introduction. *Journal of Language and Social Psychology*, 18(1), 7.

Preston, D., Ito, R., Evans, B., & Jones, J. (2000). Change on top of Change: Social and Regional Accommodation to the Northern Cities Chain Shift. In H. Bennis, Ryckeboer & J. Stroop (Eds.), *De Toekomst van de Variatielinguïstiek* (special issue of *Taal en Tongval* to honor Dr. Jo Daan on her ninetieth birthday ed., pp. 61-86). Amsterdam: Meertens.

Preston, D. R. (1999). *Handbook of perceptual dialectology*. Amsterdam; Philadelphia: J. Benjamins.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18(1), 10-31.

Purnell, T., & Koplin, L. (2003). Perceptual differences in source-filter characteristics of racially affiliated dialects of American English. *Journal of the Acoustical Society of America*, 113(4), 2328.

Rakerd, B., & Plichta, B. (2003). *More on perceptions of /a/ fronting*. Paper presented at the NWAV 32, University of Pennsylvania.

Rodman, R. (1999). *Computer speech technology*. Boston: Artech House.

Rothenberg, M. (1995). Pneumotachograph Mask or Mouthpiece Coupling Element for Airflow Measurement During Speech or Singing, *U.S. Patent No. 5,454,375*.

Rothenberg, M. (1999). *A New Method for the Measurement of Nasalance*. Unpublished manuscript.



- Sankoff, D., Rand, D., Rousseau, P., Hindle, D., & Pintzuk, S. (2004). GoldVarb (Version 2.1). Montreal: Centre De Recherches Mathematiques, University of Montreal.
- Sensimetrics. (1997). Hlsyn (Version 2.2). Cambridge: Sensimetrics Corporation.
- Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human Communication: A Unified View* (pp. 51-66). New York: McGraw-Hill.
- Stevens, K. N. (1985). Evidence for the Role of Acoustic Boundaries in the Perception of Speech Sounds. In V. A. Fromkin (Ed.), *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*. Orlando: Academic Press, Inc.
- Stevens, K. N. (1989). On the quantal theory of speech. *Journal of Phonetics*, 17, 3-45.
- Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, Mass.: MIT Press.
- Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6, 111-128.
- Stockwell, R., & Minkova, D. (1997). On drifts and shifts. *Studia Anglica Posnaneisia*, 31, 283-303.
- Stockwell, R., & Minkova, D. (2000). *English vowel shifts and "optimal" diphthongs: Is there a logical link?* Paper presented at the Optimal approaches to language change, Georgetown University, Washington, D.C.
- Strand, E., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbob (Ed.), *Natural language processing in and speech technology: Results of the 3rd KONVENS conference, Bielefeld, October 1996*. Berlin: Mouton.
- Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086-1100.

- Thomas, E. (2000). Applying phonetic methods to language variation. *American Speech*, 75, 368-370.
- Thomas, E. (2002). Sociophonetic Applications of Speech Perception Experiments. *American Speech*, 77, 115-147.
- Williams, F. (1976). *Explorations of the linguistic attitudes of teachers*. Rowley: Newbury House Publishers INC.
- Wright, J. T. (1986). The behavior of nasalized vowels in the perceptual vowel space. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 45–67). Orlando, FL: Academic Press.
- Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, 68, 1523-1525.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02736 7279