SOCIOPHONETICS

AN INTRODUCTION

ERIK R. THOMAS

SOCIOPHONETICS

Also by Erik R. Thomas

AN ACOUSTIC ANALYSIS OF VOWEL VARIATION IN NEW WORLD ENGLISH

THE DEVELOPMENT OF AFRICAN AMERICAN ENGLISH (*with Walt Wolfram*)

AFRICAN AMERICAN ENGLISH SPEAKERS AND THEIR PARTICIPATION IN LOCAL SOUND CHANGES: A COMPARATIVE STUDY (*ed. with Malcah Yaeger-Dror*)

Sociophonetics

An Introduction

Erik R. Thomas

palgrave
macmillan

First published 2011 by
PALGRAVE MACMILLAN

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

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

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

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

ISBN 978-0-230-22455-1 hardback

ISBN 978-0-230-22456-8 ISBN 978-1-137-28561-4 (eBook)
DOI 10.1007/978-1-137-28561-4

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

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

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

To my four mentors

Guy Bailey
Ronald R. Butters
Robert D. King
Walt Wolfram
Without any one of them, I wouldn't be here to write this book.

Brief Contents

Preface xiii

1 The Place of Sociophonetics 1
2 Production 17
3 Perception 55
4 Consonants 90
5 Vowels 138
6 Prosody 184
7 Voice Quality 224
8 Combinations of Different Types of Variables 251
9 Variation and the Cognitive Processing of Sounds 256
10 Sound Change 273
11 Social Factors and Phonetics 290
12 Lateral Transfer 303

Notes 312
References 315
Index 340

Contents

Preface xiii

1 The Place of Sociophonetics **1**
1.1 A youthful field 1
1.2 What is sociophonetics? 2
1.3 History of sociophonetics 8
1.4 Plan of this book 10

2 Production **17**
2.1 Production versus perception 17
2.2 Basic acoustic and signal-processing concepts 17
2.3 The source-filter theory 30
2.4 Measuring the fundamental frequency 37
2.5 Measuring formants 41
2.6 Measuring articulation directly 50
2.7 Brain imaging techniques and production 52

3 Perception **55**
3.1 Approaching speech perception 55
3.2 Auditory transformation 56
3.3 Perception experiments 61
3.4 Brain imaging techniques and perception 87

4 Consonants **90**
4.1 Lots of attention, but ... 90
4.2 Manner of articulation 90
4.3 Place of articulation: transitions 98
4.4 Place of articulation: frication and burst spectra 107
4.5 Place of articulation: direct measurement of articulation 115
4.6 Voice onset time 116
4.7 Processes affecting coda stops 119
4.8 Other voicing properties 125
4.9 Laterals 126
4.10 Rhotics 129
4.11 Deletion of consonants 135

5 Vowels **138**
5.1 Vowels in context 138
5.2 Duration 139

5.3 Applying vowel formant analysis 145
5.4 Points to measure 150
5.5 Plotting vowel formant measurements 154
5.6 Vowel normalization 160
5.7 Interactions between vowel variation and voice quality 171
5.8 Steady-state patterns 172
5.9 Undershoot 174

6 Prosody 184
6.1 Diverse and full of opportunities 184
6.2 Global factors 185
6.3 Lexical prosody 189
6.4 Intonation 200
6.5 Other post-lexical prosody 218

7 Voice Quality 224
7.1 The status of voice quality 224
7.2 Measurement domain 225
7.3 Laryngeal factors 226
7.4 Supralaryngeal settings 239
7.5 Tenseness 248

8 Combinations of Different Types of Variables 251
8.1 Combining variables in production studies 253
8.2 Combining variables in perception studies 254
8.3 Links between sociophonetics and abstract linguistic issues 255

9 Variation and the Cognitive Processing of Sounds 256
9.1 Phonetic processes aren't always automatic 256
9.2 Variability in cues used for contrasts 257
9.3 Phasing of gestures 259
9.4 Variations in undershoot 260
9.5 Prototype Theory and Exemplar Theory 262

10 Sound Change 273
10.1 (In)separability of linguistic and social factors 273
10.2 Approaches to the causes of sound change from phonetics 273
10.3 Approaches to the causes of sound change from sociolinguistics 279

11 Social Factors and Phonetics 290
11.1 Socio ... linguistics 290
11.2 Intraspeaker variation 292
11.3 Clustering of subjects 294
11.4 Cognitive sociolinguistics 300

12 Lateral Transfer **303**

Notes 312
References 315
Index 340

Preface

This introduction to sociolinguistics grew out of my long interest in the interface between sociolinguistics and phonetics. There's never been a time when I didn't see the two as supremely compatible and intertwined. This interrelationship holds on numerous levels, from the most concrete to the most abstract. I find it unfortunate that sociophonetics is too often seen as merely methodological. Its theoretical aspects, discussed in the latter chapters of this book leading up to the lateral transfer model in Chapter 12, can help sociolinguistics to become more integrated into linguistics as a whole, and indeed to be seen as central, as it should be.

This book is intended for students in both sociolinguistics and phonetics. Both groups can use this as a guide to phonetic techniques and to some theoretical issues that affect both fields. It's designed differently from any previous phonetics or sociolinguistics textbook. Most of all, though, I hope that it aids both groups in seeing how interrelated their fields are.

The approach I take in this book is that there's no one right method for performing an analysis. Instead, there are usually several ways, each with its advantages and disadvantages. It's better to make an informed decision among different techniques than to stick blindly with one. Moreover, the equipment and programs used to conduct analyses should be largely interchangeable. Hence, the methods presented in this book can be implemented using a variety of analysis tools. Nevertheless, to create the figures and to extract data for some of the tables, I relied primarily on Praat, version 5.1.04 (Boersma and Weenink 2009), for acoustic analysis, and Origin™, version 6.1, for plotting. Figure 7.10 is a screenshot.

Many of the spectrograms and other acoustic analyses are of my voice or of my daughter's. These recordings were made with a Marantz Portable Cassette Recorder PMD222, using a Sony F-V200 omnidirectional microphone, and were recorded in a soundproof booth (or sometimes in a quiet room) and digitized later at a sampling rate of 44.1 kHz with lowpass filtering at 20 kHz. Some of the recordings of North Carolinians were also made in a soundproof booth but were recorded digitally with a sampling rate of 22.05 kHz and lowpass filtering at 10 kHz. The remaining acoustic analyses were based on field recordings, usually made with a Marantz Portable Cassette Recorder PMD222 and a Sony F-V200 omnidirectional microphone. They were digitized at a sampling rate of 44.1 kHz with lowpass filtering at 20 kHz.

A number of people provided assistance in the production of this book, and I owe them considerable thanks. Tyler S. Kendall created the website in which the downloadable soundfiles used for the practice exercises are stored. He did most of the work in developing the program on which the demonstrations of vowel

normalization techniques in Chapter 5 were run. He also created Figure 6.1 and the prototypes for Figures 6.2 and 6.3 for me. I thank my daughter, Jane Thomas, for her cooperation and patience as I recorded her voice in a number of sessions in the soundproof booth. I couldn't ask for a nicer daughter. Warren Maguire provided me with recordings of speakers from Newcastle-upon-Tyne, England, and Northern Ireland. Margaret Maclagan, Jaclyn Ocumpaugh and Walt Wolfram provided other recordings. My mother, Mary C. Thomas, was instrumental in contacting the speakers from Ohio. Robert Bayley, Richard Cantu, Belinda Schouten and especially Yolanda Treviño provided me with access to the speakers from southern Texas. I thank all of the anonymous speakers whose voices are featured in various figures in this book for their cooperation. James Hillenbrand and Laura Colantoni helpfully and promptly answered a number of queries I sent them. Robin Dodsworth provided helpful advice on statistics and sociological analyses. Other key assistance was provided by Josh Rector and Charlotte Vaughn. Three anonymous referees provided useful advice that I incorporated into this book. The editorial staff at Palgrave Macmillan, particularly Kitty van Boxel, Kate Haines, Keith Povey and Felicity Noble, have been a pleasure to work with. Walt Wolfram provided constant moral support throughout the whole project. Finally, I especially thank my wife, Barbara Hunter, for her emotional support and encouragement and for her tolerance of the long, often late hours and preoccupation that this project led me into.

ERIK R. THOMAS

The Place of Sociophonetics

1

1.1 A youthful field

Sociophonetics has seemingly burst upon the linguistic scene in recent years. In the mid-1990s, the term *sociophonetics* was virtually unknown, used occasionally in phonetics but otherwise unrecognized. Foulkes and Docherty (2006) date its first use to a study of Québec French by Deschaies-Lafontaine (1974), but it was seldom used for the following 20 years. Now the term is heard widely in both sociolinguistic and phonetic circles. It has become a staple at conferences in both fields and even draws attention in phonology.

Sociophonetics, in its rapid growth, has expanded in various directions and hence hasn't been as easy to define as it might seem. Most basically, it is the interface of sociolinguistics and phonetics. Its phonetic aspects are usually limited to practices of modern phonetics, including acoustic and articulatory analysis and speech perception experiments. That circumscription excludes traditional impressionistic phonetic transcription, though some authorities might include it. The scope of sociolinguistic topics that it covers includes any kind of variation – stylistic, geographical, social class-based, gender-based, generational, ethnic and social clique-based. It also encompasses applications in fields such as language contact and forensic linguistics. That definition may seem broad enough to please everybody, but concepts of what sociophonetics is still differ a lot. Phoneticians tend to view it as comprising phonetic studies that involve any kind of dialectal variation, but not necessarily examining social aspects of language. Sociolinguists, conversely, often see it as including any sociolinguistic studies that utilize modern phonetic techniques, particularly acoustic analysis or speech perception experiments, but they don't consider any of the theoretical issues of concern to phoneticians to be essential to it. Furthermore, many people from both fields often view sociophonetics as a methodological construct, whereas others recognize a theoretical side to it as well. Within sociolinguistics it has become identified with the study of vowel shifting, though this represents only one corner in a roomful of possibilities.

Such different viewpoints on definition are the mark of a field that is just emerging. As widespread as sociophonetics is becoming, though, the time

has come for the field to begin to define itself. Among the aims of this book are those of tying together the various strands – sociolinguistic and phonetic, methodological and theoretical – and of expanding the purview of the field by showing readers the range of possible issues that sociophonetics can address. Most of all, I'd like readers to see why the research aims of sociolinguistics and phonetics aren't as disparate as they seem and how they actually converge on the goal of understanding the cognitive and diachronic aspects of language.

1.2 What is sociophonetics?

So what's sociophonetics all about? To begin with, it's intrinsically empirical. Its theoretical aims are closely linked to its empirical methods. That is, it views language as something that must be observed to be understood, and it generally holds that hypotheses should be accompanied by data collection. Phonetics is the aspect of language that is most easily observed because it can be physically measured in sundry ways, and sociophonetics takes phonetics as an entry point into language. Yet sociophonetics also views variation and change as the most fundamental of properties of language. Speakers adjust to their environment by adjusting their phonetics. Phonetic properties provide speakers with more parameters to vary than other realms of language (with the possible exception of lexicon, but most lexical variables appear infrequently in speech). Hence sociophonetics holds that an understanding of the cognitive forces underlying speech cannot be based on a notion of language as static. Instead, it has to be based on a perspective of language as inherently unstable, allowing speakers to adapt and accommodate to social situations in which they find themselves.

Docherty et al. (1997) discuss the tension between two well-known approaches to linguistic research. One, the 'top-down' approach, begins with a theory developed by the researcher based on a small amount of data. The theory is intended to provide an elegant means of accounting for the data. The other approach, 'bottom-up', starts with a survey that yields a large amount of data, after which researchers construct a hypothesis that best matches the data. The hypothesis isn't always perfectly elegant because large corpora of data almost always produce a few loose ends that seem to fit no pattern, especially where human behaviour is concerned. Formal fields of linguistics, including phonology, have customarily favoured top-down designs, though in recent years the laboratory phonology movement has emphasized more bottom-up approaches. Sociolinguists, on the other hand, have privileged bottom-up procedures because of their emphasis on fieldwork, survey designs and interviewing techniques. Docherty et al. (1997) take, as an example of the value of bottom-up approaches, one variable, the glottalization of voiceless stops in the English of Tyneside in northern England. Previous studies had formulated hypotheses about the distribution and phonological specification of glottalization using a top-down perspective. Docherty et al., however, showed that a survey of Tyneside English produced data that contradicted the expectations of the top-down hypotheses. Furthermore, it yielded other unexpected results.

For example, glottalization was disfavoured in Tyneside English before pauses, a configuration at odds with patterns found in other dialects.

Like sociolinguists, experimental phoneticians prefer 'bottom-up', empirical routes to hypotheses. The two groups differ on methodological priorities, however. Sociolinguists usually place a high priority on defining the community that they're studying. The 'speech community' circumscribes a sociolinguistic study, and the history and social structure of the community under study factor heavily into both the hypotheses being tested and the findings. Phoneticians are seldom concerned about speech communities except in the broadest terms – e.g. speakers from a particular nation who speak the same language. Sociolinguists tend to be far more concerned than phoneticians with sampling techniques. They place a lot of emphasis on finding 'typical' or 'representative' speakers and with population sampling techniques as a whole. Even though they rarely attempt random samples and often rely on convenience sampling, they frequently construct stratified samples in which the possible combinations of social classes, age groups, sexes, ethnic groups and other factors are represented. At other times, they target specific groups, such as members of a social clique or of a minority group. Phoneticians sometimes target particular groups, such as bilinguals with a particular L1 and L2 (where L1 is 'first language' and L2 is 'second language') or speakers with hearing impairments, but most of the time they don't concern themselves with population sampling. Their subjects are frequently whatever students or colleagues they can persuade to take part in their studies. On the other hand, phoneticians often try to get large amounts of data from individual subjects, which reduces potential random errors. Sociolinguists, who like to get larger samples of subjects, often aren't as concerned with getting large samples of data from individual subjects, except when they target stylistic variation.

Another key difference is that sociolinguists aim for naturalness in speech samples, while phoneticians aim for experimental control. That is, sociolinguists value speech samples that are closest to conversations in everyday encounters. They attempt to minimize the 'observer's paradox' or 'Hawthorne effect', the fact that subjects' behaviour changes when they know that they are being observed. This isn't to say that sociolinguists always obtain natural conversations – in fact, interview-style conversation is probably most common, and they often elicit read speech – but speaking style is a critical issue for them. Phoneticians, conversely, value replicability most highly. When conducting an experiment, it's important to them to control for as many factors as possible, both to isolate the factors that are under study and to make the study repeatable to see whether the same result is obtained. Hence, most of the time they use read speech that has been recorded in a laboratory when they study speech production, or they conduct laboratory experiments to study speech perception. The sociolinguistic need to capture naturalistic situations can preclude replication of studies. At the same time, phoneticians' need for controlled experimentation tends to preclude naturalness. The empirical challenge of sociophonetics is to take the methodological concerns of both traditions seriously.

While the empirical practices of sociophonetics require the balancing of sociolinguistic and phonetic priorities, the research questions that it addresses

must be related to the larger aims of linguistics as a whole. Linguistics, over its history, has been concerned primarily with two questions. Each of these questions has been modified from its earlier form over the years. The first question is how and why language changes. Linguistics has traditionally focused more on the intricacies of sound change than on other types of linguistic changes. It was the subject of the first great controversy in linguistics, the dispute over the Neogrammarian Hypothesis during the 1870s and 1880s. The Neogrammarian Hypothesis stated that sound changes were exceptionless and that they could be conditioned only by phonetic factors (Osthoff and Brugmann 1967). The main objections to it over the past century have come from the question of whether sound changes can be conditioned by morphology and syntax and from adherents of 'lexical diffusion', the notion that sound changes spread through the lexicon word by word (see Wang 1977). Another development came with the advent of phonetics, when it became clear that sounds are not discrete entities but grade into each other. If anything, this development made it easier to see how sounds could change. During the twentieth century, the study of linguistic change gradually shifted towards examining language contact instead of just genetic developments as scholars such as Uriel Weinreich (1953) drew attention to the importance of contact. The major shift in the question of how and why language changes, however, was associated with Labovian sociolinguistics. Labov (1975) popularized 'the use of the present to understand the past' in linguistic change. This approach is an instantiation of the uniformitarian principle, a widely recognized principle used in biology, geology, physics and other sciences, which asserts that processes observable today are the same processes that have always operated. As it applies to linguistic change, it involves examining how linguistic innovations originate and are propagated within speech communities. In effect, the study of how linguistic change occurs has been equated in part with sociolinguistics. Yet the aim of sociolinguistics is to study how and why language *varies*. The result, and a productive result it has been, is that variation is seen as inseparable from change and vice versa. Thus, this first major question in linguistics can be reformulated as: how and why does language vary and change?

The other major concern of linguistics is that of how language is structured. This question, formulated in terms of language as an abstract object, was essentially the prime concern of Saussure, the American Structuralists and the Prague School. Language had to be objectified because of the rudimentary state of knowledge at that time about the neural structures underlying language. Bloomfield (1933: 34–6), for example, describes what was then known about aphasias, but otherwise avoids discussion of the brain. Bloomfield was, by his own account, opposed to a 'mentalistic' approach to language, but by that he didn't mean that the psychological organization of language was unimportant. Instead, he argued against any 'non-physical factor' (Ibid.:32) governing language. At any rate, the neurological basis of linguistics was too little-known in his day for him to address it in any depth. Hence he and other structuralists concerned themselves with, as their name suggests, structural aspects of language, especially in phonology and morphology. Across the Atlantic, the Prague School Linguists didn't ignore the neurological basis of language.

Roman Jakobson, for example, wrote extensively about aphasias, beginning as early as 1941 (Jakobson 1962). However, they were also largely confined to discussing linguistic structures, not neural substrates of language.

The shift, though, came with Chomsky, who put the 'mind' and/or brain at the front and centre. As is well known, he has maintained that there is a special cognitive module predisposed to language. The rest of linguistics came to agree with him that the human brain is predisposed to language, and today the main point of contention is the extent of the innate components of language. The fact that this question has become so important reflects how Chomsky shifted the second major issue in linguistics. Chomsky (1965: 4) contended that 'in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior'. This mental reality was his concept of *competence*. Discovering what competence consists of became and has remained a primary focus of linguistics. More recently, Chomsky (1988: 3) formulated what he considered to be the major questions for linguistics, and the first two deal directly with the cognitive substrate of language:

1. What is the system of knowledge? What is in the mind/brain of the speaker of English or Spanish or Japanese?
2. How does this system of knowledge arise in the mind/brain?

While not all linguists attack the problem of the mental/neural organization of language the same way that Chomsky does, and not all share Chomsky's views of how language is represented psychologically, virtually all linguists now agree that understanding the mental/neural organization of language is a fundamental aim of linguistics and a large part of the field is devoted to studying it. This second general issue in linguistics has, then, been shifted from how is language structured? to how is language structured in the mind/brain?

The two great concerns of linguistics seem disparate, but there are links between them. The most obvious link is that any change or variation in a linguistic characteristic necessarily entails a change or variation in the internalized grammar. Hence, the study of linguistic variation is also the study of neurolinguistic variation. In addition, though, language variation and change can offer other cues into the grammatical and neurological structure of language. Linguistic changes may reflect grammatical parameters: for example, sound changes normally involve one sound changing into a similar sound. Proposed derivations in phonological theories frequently follow historical processes closely, famously (or notoriously, depending on one's point of view) the derivations involving long/short vowel alternations in English in Chomsky and Halle (1968). Variation, whether intraspeaker, intralinguistic or interlinguistic, may show that certain linguistic factors are internalized and aren't automatic consequences of the linguistic processing system or the articulators. Kingston and Diehl (1994) illustrate that fact in discussing cross-linguistic differences in the phonetic cues used to effect the [±voice] phonological feature. Finally, variation can indicate that particular facets of language not otherwise considered part of the grammar are internalized in some way. Hymes (1974: 174 ff.), for example, proposed that linguistic competence shouldn't just encompass

referential aspects of language, but should also contain the pragmatic knowledge a speaker has about what variants mean socially and how to use them for stylistic purposes.

The division between linguists who study variation and change in language and those who study structure and neural substrates is sometimes broken down into a division between 'functional linguistics' and 'structural linguistics'. Functional linguistics examines not just variation and change but everything that Ferdinand de Saussure (1983) had dubbed *parole*, that is, the social uses of language. Structural linguistics has its roots in Saussure's concept of *langue*, the grammatical system of language, though it now operates through the lens of Chomsky's *competence*, which has superseded *langue* and which, unlike *langue*, is seen as speaker-internal. The functional/structural division has been recognized widely within linguistics. For example, Hymes (1974) devotes a chapter to the division, and an entire two-volume compendium (Darnell et al. 1999) has appeared with the division as its theme. Under this classification, the generative fields – phonology, morphology, syntax and semantics – as well as neurolinguistics, much of language acquisition study and certain aspects of phonetics, are structural. Sociolinguistics, historical linguistics, pragmatics, anthropological linguistics and many areas within phonetics are functional.

I find this structural/functional division rather unfortunate. First, the grammatical basis of language has not proved as isolatable as is often assumed in structural linguistics. Chomsky (1980) divided the human language capacity into three 'mental organs': the grammar, the conceptual system (lexicon and semantics) and pragmatic competence. The last of the three, as Hymes (1974) argues at length, is a key concern of sociolinguistics. Neuroscience, however, has not found these functions easy to separate within the brain. They may well be too highly interconnected to untangle. Current understandings that brain function is as network oriented as modular make one wonder whether we should be trying to isolate the different linguistic capacities. The relationship between phonology/phonetics and pragmatic competence is likely to be especially close because so many phonetic and phonological variables in any language show stylistic variation. In English, for example, the present participial and gerundive suffix is [ɪn] in less formal styles and [ɪŋ] in more formal ones; rising final intonational contours may signal deference; and coarticulation between vowels and approximants is strongly influenced by rate of speech, which in turn is affected by the social setting. Phonetics and sociolinguistics, and hence sociophonetics, can thereby have quite a lot to contribute to what has been the domain of structural linguistics.

Second, fields classified as functional, including both sociolinguistics and phonetics, are the seat of considerable research relevant to language structure. The phonetic details that phoneticians concern themselves with and the lectal and stylistic variants that sociolinguists deal with thus take on new meaning. They are vital to understanding phonological competence.

The functional versus structural issue has been a concern not just in linguistics as a whole but within sociolinguistics too. Hymes (1974) divides sociolinguistics into three areas of interest. The first, which he calls 'the social as well as the linguistic', he identifies as real-world applications of sociolinguistics,

such as in education and law. The other two are theoretical in nature. The second area, 'socially realistic linguistics', is essentially Labovian or quantitative sociolinguistics, and he (1974: 196) identifies it with the traditional linguistic concerns of 'the nature of linguistic rules, the nature of sound change'. The third area, 'socially constituted linguistics', is, he says (ibid.), 'concerned with social as well as referential meaning'. I would assert that the second area is the basic focus of sociophonetics. Although in linguistics as a whole the functional versus structural opposition, I think, obscures aspects of linguistic organization, within sociolinguistics it takes on distinctly different connotations. The structural outlook maintains language as the focus of study, and, when Hymes speaks of what he recognizes as a second area of sociolinguistics, 'the nature of linguistic rules, the nature of sound change', he is speaking of the two basic concerns of linguistics, respectively: how language is structured (whether in the mind/brain or not) and why language varies and changes. On the other hand, the functional outlook, represented by Hymes's 'socially constituted linguistics', places its focus on social meaning. That is, discovering language is no longer the final aim, but instead language, as an index of social meaning, becomes a way of discovering social function. Hence Hymes's third area of sociolinguistics addresses a question that is, in its essence, sociological. Trudgill (1978) makes a similar point in noting that some studies labelled as sociolinguistics address sociological questions, some address purely linguistic questions, and some lie on the border. Sociophonetic methods can offer linguistic sociologists useful tools for determining how language indexes social meaning. In fact, Chapter 11 will touch on some ways sociophonetic methods can do so, and specific applications will appear in the text and practice exercises of Chapters 4–7 as well. However, it's unfortunate that the field of sociology has never developed a linguistic branch. Where can linguistic sociologists go? In addition to the ethnography of speaking, there is Joshua Fishman's sociology of language movement and the *International Journal of the Sociology of Language*, but most of its contributors aren't in sociology departments. Linguistic sociologists are forced either to melt into linguistic anthropology or to try to fit into linguistics. The incongruity hinders and diminishes their labour and insights, which address an essential aspect of human life.

More recently, Eckert (2005) has taken up the status of functional approaches within sociolinguistics. Eckert, as demonstrated by her previous work (e.g. Eckert 1989a), resides partly in the Hymesian tradition of focusing on social meaning. From her perspective, she sees variationist work as representing three waves. The first wave, typified by Labov's (1966) survey of the Lower East Side of Manhattan, examined correlations between linguistic variation and broad demographic characteristics. The second wave, exemplified by Milroy's (1980) study of Belfast, took an ethnographic approach, examining, for example, how linguistic variation is correlated with an individual's engagement in local community networks. The third wave, according to Eckert, focuses on stylistic variation and how speakers use it to project different identities. Sociophonetics takes stylistic variation seriously. In contrast to the characterization of formal linguistics as treating styles 'solely as successive modifications of an ordinary grammar' (Hymes 1974: 177), sociophonetics treats styles as integral parts

of how language is internalized. It also recognizes the intertwined nature of linguistic variation and social meaning. However, it offers an alternative to having the discovery of social function as the ultimate goal of variation studies. Instead of focusing on how language variation and the construction of speaking styles are a means of constructing social meaning, sociophonetics addresses how speaking styles, with their inherent social meanings, are a path to understanding how language is structured. It thereby views social meaning as a crucial aspect of the cognition of language. Some recent sociophonetic papers illustrate this focus. For instance, Foulkes and Docherty (2006) discuss how language variation can inform Exemplar Theory, a new and controversial conception of how language acquisition and the cognitive organization of language take place. Similarly, Purnell et al. (2005b) explore the plasticity of a phonological contrast by examining changes in the cues used to produce the voicing contrast across four generations of residents of a German–American community in Wisconsin.

The notion that sociolinguistics can provide clues to how language is internalized is by no means new. During the heyday of generative phonology, sociolinguists developed variable rules that represented the ordering of constraints on phonological variables (e.g. Labov 1969; Cedergren and Sankoff 1974). The rules were construed as reflecting internalized processes that governed the occurrence of variants. In the years since the popularity of variable rules has waned, other papers relating sociolinguistic findings to the structure of internalized grammar have appeared sporadically. Some of these efforts are framed in terms of more recent phonological theories, such as Optimality Theory (Nagy and Reynolds 1997). Optimality Theory holds that phonologies of languages differ in the importance that each places on various competing constraints. Other efforts aren't tied to a particular phonological theory. For example, Labov (e.g. Labov 1994) has proposed a phonological feature [±peripheral] to account for why, over time, long or tense vowels tend to be made higher in the mouth and short or lax vowels lower in the mouth. Sociophonetics represents the continuation of the movement to relate language variation to the cognitive structure of language. As such, it is also the continuation of the side of sociolinguistics that is focused on the two traditional questions of linguistics: how language is structured in the mind/brain and how/why language varies and changes. Moreover, no subfield of linguistics is better positioned to show how those two questions are interrelated.

1.3 History of sociophonetics

Sociophonetics began before the name was apparently coined in 1974. This precedence holds even if we exclude studies that used only impressionistic transcription, since, as I mentioned earlier, *sociophonetics* is usually used to denote approaches using more modern phonetic techniques. Within sociolinguistics, the genesis of sociophonetics is, basically, Labov, Yaeger and Steiner (1972), though even earlier Labov (1963) made limited use of acoustic data. Labov et al. (1972) ushered into sociolinguistics the use of acoustic analysis

for studying vowel variation. Their primary aim was to determine the principles governing vowel shifting. The equipment available in those days was a lot less user-friendly than the equipment we have today. They had to print out large numbers of spectrograms and measure the formants by hand from the printed spectrograms. Today, spectrographic software and linear predictive coding (LPC) (Atal and Hanauer 1971) have made measurement of formants, the main parameters of vowels, paperless and much faster. However, reliance on vowel formant measurements and on plots showing the first two formants has remained standard for studying vowel variation.

Labov and his students at the University of Pennsylvania were nearly the only sociolinguists to use acoustic analysis of vowels in dialectal studies for two decades. Besides Labov et al. (1972), Hindle (1980), Labov (1980, 1991), Ash (1988) and Veatch (1991) represent works by Labov's research team during this earlier stage. Among the few outsiders to conduct acoustic research on dialects of English were Habick (1980), Maclagan (1982) and Godinez (1984). The situation changed during the 1990s, however, as numerous researchers beyond the University of Pennsylvania adopted spectrographic methods for vowel analysis. Most of these studies – e.g. Thomas and Bailey (1992); Esling and Warkentyne (1993); Ito and Preston (1998); Fought (1999); Wolfram et al. (1999); Fridland (2000, 2003); and Thomas (2001) – have been conducted in North America. However, a growing number – e.g. McClure (1995); Cox (1999); Watt and Tillotson (2001); Fabricius (2002); Deterding (2003); Torgersen and Kerswill (2004); and Sharbawi (2006), in addition to Maclagan's (1982) previous study of Australian English – have taken place elsewhere and analysed varieties of English from other parts of the world. The flowering of acoustic analysis of vowel variation represents an encouraging democratization of sociophonetics. However, acoustic vowel variation work has barely spread beyond analysis of English. Among the few exceptions are studies of Dutch vowel variation by van Heuven et al. (2002), Adank et al. (2007) and Jacobi et al. (2007).

Within phonetics, sociophonetics hasn't had as clear of a beginning as in sociolinguistics. Over the years, occasional studies appeared that utilized dialectal or other variations, e.g. Fourakis and Port (1986), Henton (1988) and Munro et al. (1999), but sociophonetics was usually neglected. Foulkes and Docherty (1999: 22) asserted that phoneticians were 'treating variation as a nuisance', as something that represented an obstacle to isolating other factors that they were interested in. Now, though, it's clear that phoneticians are taking notice of variation. This interest is most obviously reflected at phonetics conferences. The International Congress of Phonetic Sciences now regularly holds sessions devoted to sociophonetics. Meetings of the Acoustical Society of America usually have a number of sociophonetic presentations and occasionally a session, and even the International Congress on Speech and Language Processing has a few papers involving variation. Laboratory phonology has followed suit: the 9th Conference on Laboratory Phonology, held in 2004, had as its theme 'Change in Phonology' (see Cole and Hualde 2007).

Acoustic studies of variation in elements of language besides vowels have proceeded at a far slower pace. Docherty and Foulkes (1999) and Purnell et al. (2005a) both bemoan the shortage of acoustic studies of consonantal variation,

though both provide fine examples of how to approach this acoustically. Both demonstrate that acoustic analysis can reveal details of consonantal variation that are difficult or impossible to gauge using impressionistic analysis. Nevertheless, because sociolinguists have generally considered impressionistic analysis adequate – not to mention quicker – they have held fast to it and no sustained tradition of acoustic analysis of consonants has developed. Perhaps that will change.

The treatment of prosodic variation has been something of a joint effort among sociolinguists, phoneticians and phonologists. Acoustic studies of intonational variation got off to a slow start, since impressionistic transcription of intonation was the rule for decades. The development of the Tone and Break Index (ToBI) transcription system (Beckman and Hirschberg 1994) encouraged greater use of acoustic methods, mainly pitch tracking; even though it's an impressionistic transcription system, implementing it involves reference to pitch tracks. Work on intonational variation employing pitch tracks, generally relying on ToBI or similar systems, has expanded in recent years, and much of the work has occurred in Europe (e.g. Gussenhoven and van der Vliet 1999; Grabe et al. 2000; Selting 2003). Another aspect of prosody – prosodic rhythm – has also attracted attention recently. Prosodic rhythm is conventionally regarded as the degree of syllable-timing versus stress-timing that a language exhibits. A number of dialectal studies of prosodic rhythm, using new methods of quantitative, acoustic-based analysis, have appeared (e.g. Low et al. 2000; Deterding 2001; Thomas and Carter 2006). Other aspects of prosody, such as lexical tones and stress realization, have received less sustained attention.

Studies of speech perception as it relates to language variation have proceeded in fits and starts, but on a wide variety of research issues. Studies of dialect identification go back at least as far as Dickens and Sawyer (1952), and some of the recent studies, such as Bezooijen and Gooskens (1999) and Clopper and Pisoni (2007), have become quite sophisticated in both phonetic techniques and statistical analyses. Studies of attitudes towards different forms of speech also have a long history, as exemplified by the matched-guise experiment in Lambert et al. (1960). Synthetic manipulation of stimuli, introduced by Brown et al. (1972), has provided a means of controlling stimuli in attitude experiments. The perception of segments that are phonologically merged for some speakers but not for others has drawn some attention (e.g. Labov et al. 1991). Various other issues, such as the intelligibility of dialectal variants of segments (e.g. Labov and Ash 1997), have also been addressed in different experiments. Sociophonetic study of perception has lacked a specific focus but has figured prominently in the development of sociophonetics. See Thomas (2002b) for a detailed discussion of the history of socioperceptual studies.

1.4 Plan of this book

This book is designed to give you two kinds of information. One is specific information on how to analyse speech sounds and conduct experiments. It is intended for anyone who's beginning to conduct phonetic research or who has

experience in one type of phonetic research but would like to expand his or her areas of work. It helps to have an introductory phonetics course first, but I've written it so that you can plunge into phonetic analysis with little more than an introduction to linguistics course and perhaps a general sociolinguistics class. For many people, it's easier to learn through experience with analysis than through listening to lectures.

The other kind of information is a more abstract one about sociophonetics. We've just discussed what sociophonetics is and how it came about. In the last few chapters, we'll return to theoretical issues and discuss some that sociophonetics addresses. The chapters vary considerably in length. Some eyebrows may be raised at the length variation, but I feel it's more important for the chapters to cover topics of equal rank than for them to have a uniform length.

Before you start the subsequent chapters, you should make sure you're familiar with how to formulate and test a hypothesis. You can do exploratory and some kinds of descriptive work without a hypothesis, but most of the analyses you'll do will depend on hypothesis testing. The general procedure is shown in Figure 1.1. First, you have to figure out what you want to show in your experiment. Then you decide what your dependent variable is – normally the linguistic variable – and what the independent variable(s) are – that is, whatever factors, linguistic, social or other – that might influence the dependent variables. Next, you decide on a hypothesis. For example, you may wish to hypothesize that short durations cause vowel undershoot or that listeners can distinguish speakers from a certain region. The *null hypothesis* is the opposite

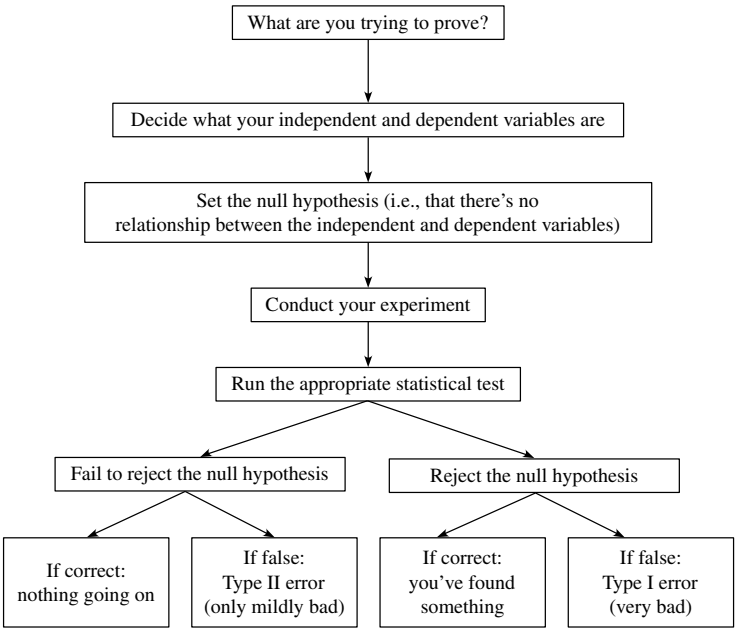


Figure 1.1 Steps in hypothesis testing.

of your hypothesis. That is, the null hypothesis states that there's no correlation between the independent and dependent variables at all. After you run your experiment and apply the right kind of statistical analysis to it, your results will either reject or not reject the null hypothesis. If the results don't reject the null hypothesis, it might mean that there's no relationship between the independent and dependent variables. However, it might mean that you didn't have enough data or the right kind of data to reject the null hypothesis, which is called a Type II or β error. If the results reject the null hypothesis, it probably means that there's a correlation between the independent and dependent variables, which is what you want. However, it's possible that there's no correlation but that, by chance, the data you have were skewed enough that they fooled the statistical test into indicating that there was. This situation is called a Type I or α error. Type I errors are considered more damaging than Type II errors. Statistical tests are designed to tell you what the chances are that their results are erroneous. They tell you this through the p (probability) value. A p value of .05 means that there's a 5 per cent chance of erroneous results, and a p value of .01 means that there's only a 1 per cent chance. In linguistics, the standard for reporting the results is $p < .05$.

I might add, though, that the statistical test is only as good as your data sample. You have to make sure that your sample is representative of the population you are studying. For some things, such as samples of phonetic data from a single speaker, the sampling isn't always an issue, though it can be insofar as speaking style is concerned. When samples of people are involved, though, ensuring that the sample is representative is crucial. Sampling is touched on a little in Chapter 11. Sometimes, usually when you're studying a small group such as a social clique, you can get data from the entire population, in which case you aren't sampling at all. Otherwise you have to sample. Kalton (1983) provides a taxonomy of sampling techniques. *Random samples* are ideal for sociological research, but they're difficult to get and they usually necessitate short interviews that preclude many kinds of linguistic research. Few linguistic studies, e.g. Bailey et al. (1993), have used them. A simple random sample involves one sample of the entire population, while a *stratified* random sample involves separate samples of different groups within the population. *Quota samples*, in which a certain number of people who meet each of various combinations of features are interviewed, have been used for some linguistic studies, such as Wolfram (1969). For a quota sample, the number of subjects with each combination may or may not represent the proportion of such people in the whole population. *Judgement samples* resemble quota samples, but the researcher chooses subjects deemed representative of their respective groups. *Convenience samples* are the most commonly taken kind in linguistics, but the least desirable for statistical comparison. A convenience sample is one that includes whatever people you were able to access. When you use a convenience sample, you assume, overtly or implicitly, that your sample isn't skewed in a way that compromises your findings. As you can imagine, that assumption isn't always safe. Nevertheless, if you're studying speech production, convenience sampling often lets you get longer interviews because these samples are acquired in less impersonal ways, such as through social networks, than quota or random samples are. With

longer interviews, you'll probably be getting a more representative sample of the subject's speech than you would with the short interviews that random sampling usually involves.

I've organized this book in a somewhat modular fashion. The next two chapters discuss speech production and speech perception, respectively, and how to study them. Quite a bit of knowledge about acoustics is needed for each one, so Chapter 2 begins with a general overview of what you'll need to know about acoustics. Many techniques of acoustic analysis are covered in this chapter, rather than in later ones, because they apply to more than one kind of variable. Chapter 3 discusses techniques and experimental designs that are useful in sociophonetics for studying speech perception. The following four chapters discuss the different kinds of variables: consonants, vowels, prosody and voice quality. These have not had equal attention from sociophoneticians; vowels have attracted by far the most research. Besides showing you how past approaches work – especially in vowel study – I want to introduce you to other techniques that are useful for the lesser-studied variables and even some new techniques for vowel study. I hope that the imbalance in attention researchers have given to different kinds of variables can be evened out in the future. Most of the techniques discussed in Chapters 4–7 relate to speech production, but once you understand them it shouldn't be hard to see how you can adapt the information to experiments targeting speech perception.

The final five chapters take on a more abstract bent. In Chapter 8 I argue for the necessity of taking a broad view of variation by combining different kinds of phonetic variables in analyses, both in production and in perception. Chapter 9 brings us around to cognitive processing related to phonetics and phonology. Sociophonetics can address this issue in a number of ways, and, as I've indicated earlier, cognitive processing of language should be one of the primary aims of sociophonetics. We'll explore what sociolinguistic 'knowledge' is and how sociophonetics can be used to test the recently formulated Exemplar Theory, which relates especially strongly to phonology. Chapter 10 discusses theories of sound change. The origin, or actuation, of sound changes has vexed linguists for two centuries, and we'll examine some of the different ideas scholars have proposed to account for it. Sociophonetics has revolutionized the way sound change is examined, though the focus has been almost entirely on vowel shifting. Different approaches apply to shifting and to phonological restructurings (mergers and splits). Sociophonetics has also tied sound change study unalterably to sociological factors. Chapter 11 addresses in more detail how sociological constructs such as communities of practice and the role of individual identity as opposed to group identity figure into sociophonetic analysis. Like Chapter 9, it addresses cognitive encoding, this time of sociolinguistic 'knowledge'. Finally, Chapter 12 asserts that the traditional Lamarckian and Darwinian approach to language change and cognition is inappropriate. A new model called lateral transfer with origins in molecular studies of biological evolution works better, ties the sociolinguistic, phonetic and cognitive threads of sociophonetics together, and thrusts sociolinguistics to the centre of linguistic theory.

I've chosen to restrict this book to linguistic issues in sociophonetics. Certain other topics that are important in sociophonetic practice are omitted for various

reasons. For one, I won't discuss recording or analysis equipment in any detail, and with rare exceptions I'll avoid discussing any particular model of equipment. Equipment changes constantly and discussion of particular models quickly becomes outdated. Discussion of how to evaluate equipment and the ways that equipment can vary would require another book the length of this one to cover adequately. All I'll say here is that you should treat equipment as a serious issue in its own right and that you should pay careful attention to the specifications for any equipment that you procure. Keep in mind that the make and model of equipment is only one factor – wear and tear make a difference, and equipment cleanliness can, too. Not only that, but the environmental conditions in which you make recordings or conduct perception experiments can affect your results. Whenever possible, use the same equipment for all recordings or trials in a single study. The main equipment issues you need to attend to in conducting a recorded interview or a perception experiment are schematized in Figure 1.2. You can test your equipment in various ways. For example, you can test the frequency fidelity of a microphone or recorder with a method as simple as sounding a tuning fork or playing a tone into it. A commonly used method of comparing the amplitude characteristics of different equipment is to play the same recording into it and then to examine power spectra of the sound recorded by the equipment. Analogous methods can be employed to test earphones used for perception experiments. Don't get too obsessed with equipment, though.

Another issue that I won't delve into in any depth is statistical analysis. A limited number of statistical procedures will take care of most of your needs. Some of the most useful and widely used tests are outlined in Table 1.1. Like evaluation of equipment, statistical procedures deserve a book-length treatment

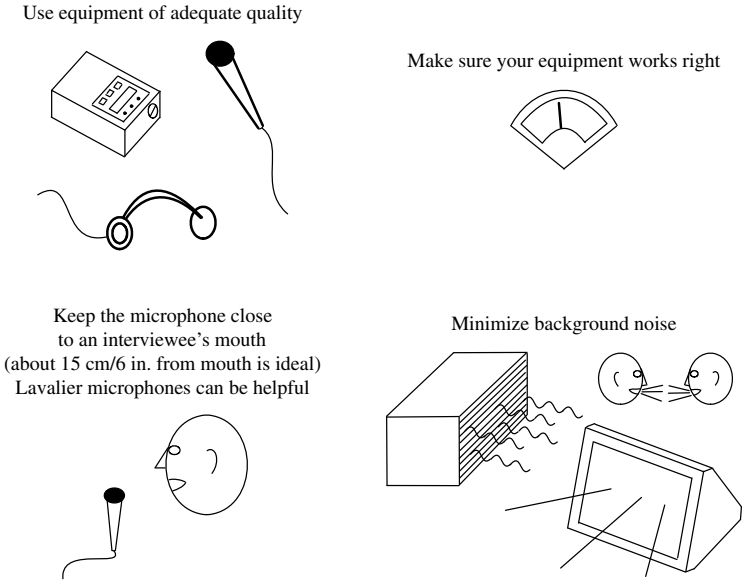


Figure 1.2 Primary equipment issues in conducting sociophonetic research.

Table 1.1 Some statistical procedures commonly used in sociophonetics.

Statistical procedure	*Use*
t-test	Testing whether sample means of two populations of continuous data are different or not (there have to be two distinct groups).
ANOVA (analysis of variance; a member of the family of tests called general linear models (GLM))	Testing whether sample means of two or more populations of continuous data are different or not; independent variables can be continuous or discrete; tells you if one differs, but doesn't tell you which one.
Post hoc tests, e.g. Tukey's *W*	After you run ANOVA, post hoc tests can show which populations differ from each other.
MANOVA (multivariate analysis of variance)	Similar to ANOVA, but used with two or more dependent variables.
Mixed models	Similar to ANOVA, but a random variable can be included.
Confidence intervals	Provides estimates of how likely it is that a sample mean matches the actual population mean, e.g. 95 or 99%; confidence intervals from different populations can be compared for overlap.
Principal components	Shows overall similarity of different individuals or groups across multiple variables.
χ^2	Used for countable data; it will tell you whether the actual distribution varies from the expected distribution, but it won't tell you which particular cell(s) differ – a post hoc test is necessary for that.
Linear regression	Used to show the relationship between two or more continuous variables; associated ANOVA shows whether they're correlated with each other, and other procedures can show the relative strength of independent variables.
Logistic regression	Used to show the relationship between two or more discrete variables.
Intraclass correlation	Used to determine the degree of difference between paired sets of data (e.g. for a reliability test).
Dendrogram	Shows relative similarity of different sets of data using a branching tree-like structure.
Multidimensional scaling	Shows relative similarity of different sets of data.
Box-and-whiskers plot (boxplot)	Provides a visual display of data distribution and reveals outliers.

themselves. In fact, several statistics textbooks designed specifically for linguistics have appeared recently: see Rietveld and van Hout (2005), Johnson (2008), Baayen (2008) and Gries (2009). Rietveld and van Hout focus on the use of ANOVA, or General Linear Models. Baayen and Gries discuss how to conduct analyses using the statistical program *R*. Johnson aims to provide a more general overview. Even these linguistically oriented statistics guides may not tell you everything you need to know for your own project. If you're using SAS™ or SPSS™ for analyses, you can often find help online. Frequently a more general statistics textbook, or one designed for social sciences, may be useful. A statistical counselling service, if your university provides it, can be invaluable.

Impressionistic phonetic transcription is excluded for reasons given earlier. It certainly has a long tradition in linguistics in general and in dialectology and sociolinguistics in particular. The International Phonetic Alphabet (IPA), of course, is the standard for phonetic transcription. Thorough discussions of IPA symbols and use of the IPA are found in the *Handbook of the International Phonetic Association* (International Phonetic Association 1999) and in Ladefoged (2001). I assume that you'll have some familiarity with the IPA and with the basic descriptions of consonants and vowels before you begin this book.

EXERCISES

1. Name several ways information on language variation can demonstrate links between variation/change in language and the organization of language in the mind/brain.
2. What are some of the ways that sociophonetic studies can meet the methodological requirements of both sociolinguistics and phonetics, especially in terms of replicability, getting naturalistic data and defining the speech community?
3. What is the status of functional and structural approaches to linguistics in general and sociolinguistics in particular? Do you agree or disagree with the perspective presented in this chapter?

FURTHER READING

Docherty, Gerard J., Paul Foulkes, James Milroy, Leslie Milroy and David Walshaw. 1997. Descriptive adequacy in phonology: a variationist perspective. *Journal of Linguistics* 33:275–310.

Foulkes, Paul and Gerard J. Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34:409–38.

Production 2

2.1 Production versus perception

Phoneticians generally divide phonetics into speech production and speech perception. Production, as you'd guess, has to do with how speakers make speech sounds, while perception has to do with how listeners process speech sounds that they hear. The means of studying production and perception differ sharply, making it logical to split phonetics instruction along the line between production and perception. Hence this chapter is devoted to introducing the basic techniques used for speech production, while Chapter 3 will cover techniques for speech perception. In both chapters there is an orientation towards methods that are useful for exploring language variation. Some techniques that are useful for only one type of sound, such as analysis of frication spectra, are discussed in later chapters.

Phonetics is sometimes split along different lines, into a three-way division among articulation, acoustics and audition. I've chosen to follow the two-way production/perception taxonomy instead because it matches up better with the way sociophonetic studies are conducted. Articulatory analyses are relatively rare in sociophonetic studies, largely because most desired information about articulation (with a few notable exceptions) can be addressed through acoustic analysis. Not only that, but the main methods of studying speech articulation directly – e.g. X-ray microbeams, magnetic resonance imaging (MRI) and devices such as electropalatographs – are poorly suited for field research. However, we'll cover several of these methods briefly in this chapter, since they can be useful for laboratory studies.

2.2 Basic acoustic and signal-processing concepts

The main tool for analysing speech production is spectrographic analysis. Spectrographic analysis gives you a visual representation of sound. This section will cover the basic terminology of acoustics and signal processing quickly. The items discussed here are also the ones that you'll need to mention when

you describe the methods for acoustic studies that you conduct. The coverage here isn't intended to go into great depth. For more detail, consult a regular phonetics textbook or a longer manual. Stevens (1998) offers an exhaustive treatment. Johnson (2003) is more concise and quite good, and other fine explanations of various elements of what I'll cover can be found in Lieberman and Blumstein (1988), Kent and Read (2002), Hewlett and Beck (2006), and Reetz and Jongman (2009).

Units of measurement

Acoustic signals have three basic dimensions. One is *time*, which is displayed in seconds (s) or milliseconds (ms). Another is *frequency*, which is measured in *hertz* (*Hz*). Older literature often refers to hertz as *cycles per second* (*cps*), which gives you a clue about what the term *hertz* means. A cycle is the period in which a sound wave goes through its pattern before repeating it. We'll see how that works shortly. The final dimension of sound is *amplitude*. The official unit of measure for amplitude is the *pascal*. Most of the time, though, you'll see amplitude measured in *decibels* (*dB*).

Decibels are a different kind of unit from what you're used to because there's no set standard for what a decibel is. In descriptions of amplitude that you read in popular publications, the 0 dB point is conventionally set to the lower threshold of human hearing – or actually the lower threshold of hearing for a young person with no injury or impairment to his or her hearing – for a tone with a frequency of 1000 Hz. Note that the dB figures you see in such publications are positive numbers. On the other hand, electronic equipment sets the reference point in other ways. Recording devices often set the 0 dB mark at the maximum level that the device can handle, so you'll see negative dB numbers listed on the recorder. Acoustic analysis software generally sets the 0 dB point to the lowest level that can be digitized, so once again the dB numbers will be positive. The decibel scale is logarithmic, not linear, unlike time and frequency. Human perception of amplitude is logarithmic, too, so the dB scale roughly reflects how our hearing works.

Sound waves

The simplest kind of sound is a pure tone – a tone at a single frequency. The waveform for this sound is represented as a sine wave, as shown in Figure 2.1. The length of the period, from start to finish, for the sine wave determines its frequency. To calculate the frequency, you measure the length of the period in milliseconds and divide it into 1000 (or measure it in seconds and divide it into 1). For example, if the period is 5 ms, the frequency of the wave is 200 Hz because $1000 \div 5 = 200$. Figure 2.2 shows sine waves with different frequencies. The representation of a waveform as a sine wave actually refers to the movement of air molecules as they transmit the sound. When a noise is created, the air molecules are perturbed. They alternately move closer together – called *compression* – and farther apart – called *rarefaction*. The bands of compressed and rarefied air molecules spread outward from the sound's point of origin, like the waves produced when you drop a rock into a pond but in three dimensions.

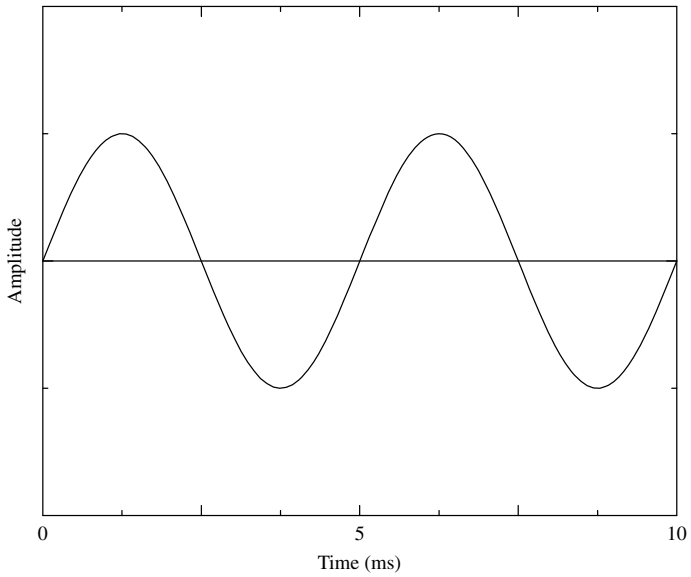


Figure 2.1 A sine waveform representing a pure tone.

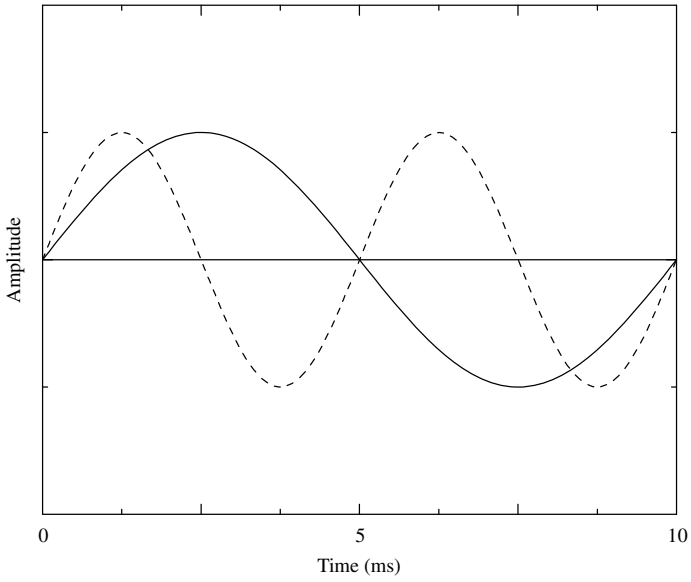


Figure 2.2 Two sine waveforms, one with a wavelength twice that of the other.

Note that compression and rarefaction are connected with how fast the air molecules move. There is an inverse relationship between the amount of compression/rarefaction and the speed the molecules are moving. It works like a playground swing. The neutral point of the swing is when the swing is hanging

straight down. When somebody is swinging, though, the swing moves back and forth, crossing that neutral point. As the swing gets closer to its most forward position, it slows down until it stops and reverses direction. Then it speeds up until it crosses the neutral point, and after that it gradually slows down until it reaches its most backward position. Air molecules transmitting a sound do the same thing. They slow down as they're compressed until they stop and begin to move farther apart. They move fastest when they're at their neutral position, and then slow down again as they become rarefied. Whereas gravity causes the swing to slow down, in sound transmission the slowing is caused by molecules bumping into each other.

Amplitude is represented by how far the peaks in the sine wave get from the neutral position. For a soft sound, the air molecules don't get compressed and rarefied very much, whereas for a loud sound they do. This difference is represented by the sine waves in Figure 2.3. The two waves have the same frequency but differ in amplitude.

Most sounds are not tones that can be represented by a single sine wave. They show more complicated waveforms and are called *complex waves*. An example is shown in Figure 2.4. However, any complex wave can be decomposed into a series of simple waves by a process called *Fourier analysis*. The waveform shown as a solid line in Figure 2.4, for example, can be broken down into the two waves shown with dashed lines, one with a frequency twice that of the other. You could also say that the two simple waves add up to make the complex wave. Note that actual speech consists of a lot more than just two components.

When Jean Baptiste Joseph Fourier first proposed the theory that any complex wave can be broken down into simple waves, he stipulated that it applied

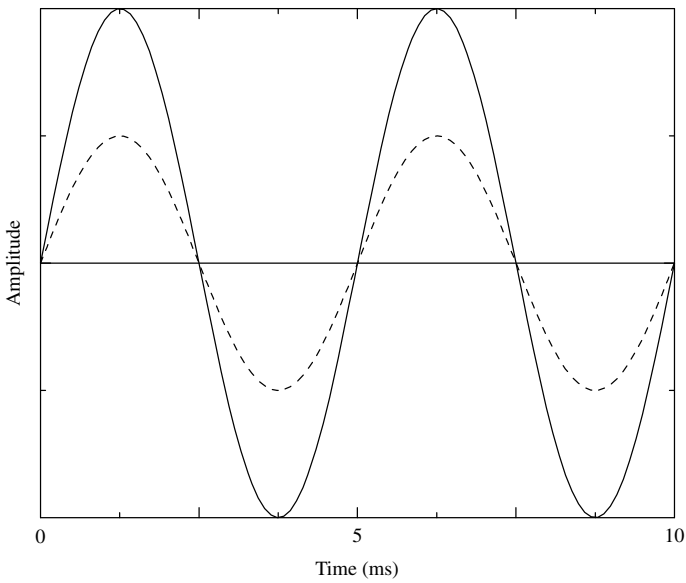


Figure 2.3 Two sine waveforms with same frequency but different amplitudes.

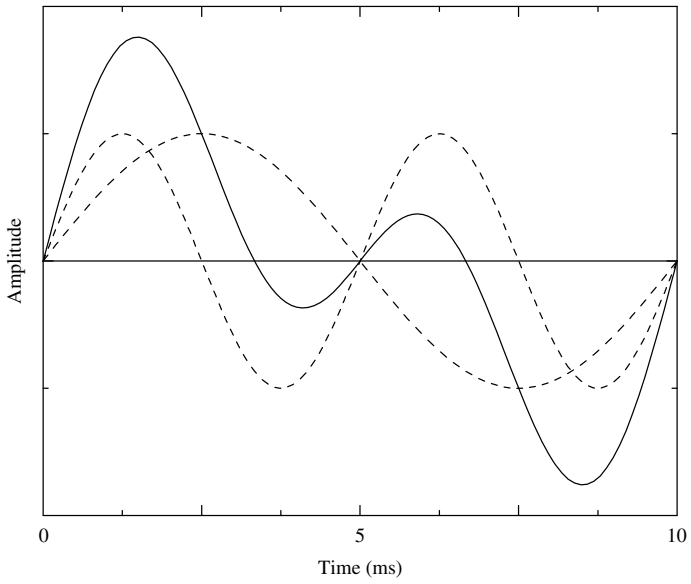
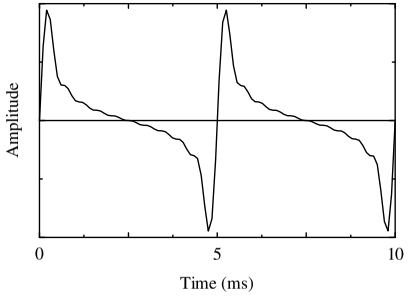


Figure 2.4 A complex waveform (sold line) and its component waveforms (dashed lines).

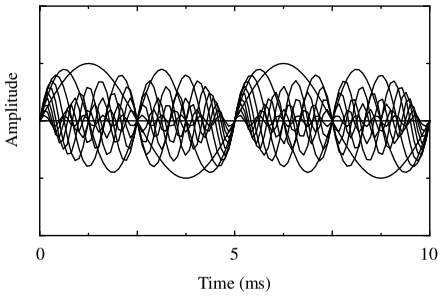
only to continuous sounds. However, in real life sounds are never indefinitely continuous. Because of that, *discrete Fourier transforms* (*DFT*) were developed. They operate by analysing only a *window* – a short segment of time – within the sound signal. The window may or may not be attenuated at each end, depending on which of various *windowing* methods is used for creating it. In addition, a special type of DFT, called the *fast Fourier transform* (*FFT*), was developed for digital computers. FFT works on digital computers because the number of points that are analysed has to be a power of two – i.e. 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, etc. As the name suggests, FFT is a lot faster than other methods of Fourier analysis.

Figure 2.5 illustrates the stages in Fourier analysis of speech. From the original signal, as in Figure 2.5(a), simple waves are extracted, starting with the lowest-frequency components. The result is a series of simple waves, as in 2.5(b). The frequency and amplitude of each wave is measured. Figure 2.5(c) gives an idealization of a spectrum of the component waves, though in real life windowing yields spectra with humps, as in 2.5(d). For voicing in speech, all the component waves – the *harmonics* – have to be at multiples of the lowest component, which is the *fundamental frequency* (F_0). The reason is that the length of the lowest wave corresponds to the time between vocal fold vibrations. Any wave with zero-crossing points at the same places as F_0 will 'fit' between the vocal pulses. All multiples of F_0 match that description.

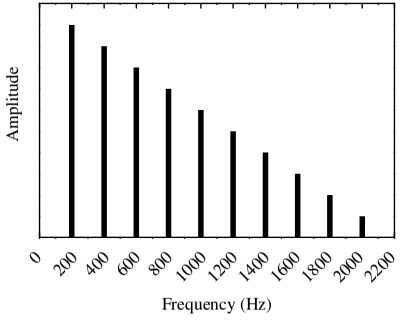
Waves can differ in another way besides frequency and amplitude. They can also differ in *phase*. Phase involves the location of starting points for waveforms relative to each other. For example, Figure 2.6 shows two waveforms that have



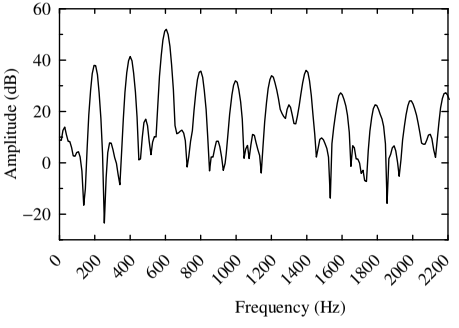
(a) Two cycles of a complex waveform that resembles speech



(b) Fourier analysis breaks the complex wave into a series of simple waves



(c) The freqencies and amplitudes of the simple waves can be represented as a series of poles in a spectrum. In speech, the component waves – the harmonics – come at multiples of the lowest-frequency wave



(d) In actual Fourier analysis of speech, analysis windows can't be infinitely narrow, so instead of poles you get a spectrum with a series of humps

Figure 2.5 Steps in Fourier analysis.

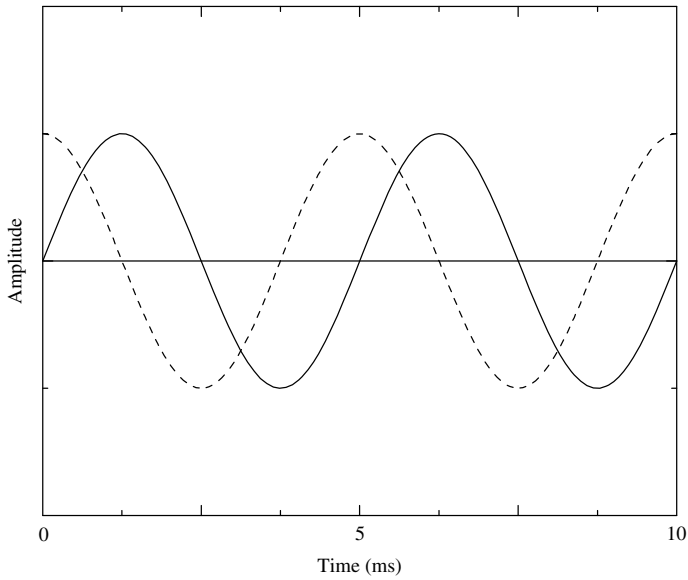


Figure 2.6 Two waveforms with the same frequency and amplitude but different phases.

the same frequency and amplitude but are out of phase with each other. Phase generally doesn't make much difference for speech sounds, with one major exception: when two sounds are in opposite phase, as in Figure 2.7. When that happens, the two sounds cancel each other out and the result is silence. Sounds in opposite phase are important for nasality, for which the oral cavity produces resonances that cancel out a resonance of the nasal cavity (for an oral consonant) or a nasal resonance cancels out an oral resonance (for a nasal vowel). These are called *antiformats* or *zeroes.* A similar phenomenon happens with laterals, for which the tongue is aligned so that it creates side-by-side cavities.

Sound signals come in two basic types: periodic and aperiodic. A *periodic* signal is one with a repeating pattern – i.e. a waveform. Bird song and the sounds of musical instruments in the wind and string groups are examples of periodic signals. In human speech, voicing, as with vowels and voiced consonants, is almost periodic. It is not perfectly periodic because the vocal fold vibration isn't repeated exactly, but it is close. An *aperiodic* signal is one in which the pattern doesn't repeat. Aperiodic signals themselves fall into two categories. A sustained aperiodic signal is called *noise.* (No, it isn't very creative, but that's the official term.) The sound of wind blowing through trees or of a fan blowing or static on a television that isn't tuned in to a channel are all noise. In speech, noise occurs in frication and aspiration. You can certainly hear the staticky quality of frication when you say a sustained [sː] sound. An aperiodic signal that isn't sustained is called a *transient.* Transients are sudden, sharp noises. Think of the sound of a knock on a door or a gunshot or what this book would sound like if you dropped it on the floor. Transients in speech

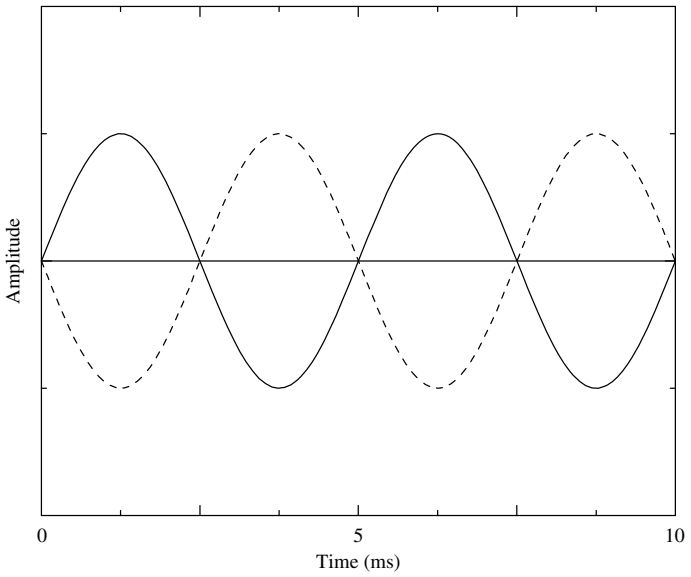


Figure 2.7 Two waveforms with the same frequency and amplitude but opposite phases – when combined, the result is silence.

are best known in one situation: when a speaker lets go of the closure for a stop consonant, there is a small pop called the *burst* when the air behind the occlusion rushes forward and hits the air in front of the occlusion. This stop burst is a transient. Transients also appear in the click sounds of the San languages of southern Africa.

Digitization

Sounds are recorded in two formats: *analogue* and *digital*. Analogue recordings, such as on cassette tapes and other older media, record a continuous signal. Their disadvantage is that they can't be copied exactly and the signal degrades over too many generations of copying. Moreover, analogue recordings aren't compatible with computers.

Computers require signals to be digitized. These days, digital recorders do this step for you, but there are digitizers and computer programs that can digitize analogue recordings. Here's how digitization works. I'm going to explain the process backwards because the reasons for the earlier steps are easier to understand this way. Digitization involves *sampling* the sound signal at intervals. That is, at each interval the energy in the waveform is measured, as shown in Figure 2.8. The result is a series of discrete measurements; some information is lost because the sampling measures the waveform at points instead of continuously. The loss is called quantization error. However, continuous measurements would produce an infinite amount of data. Sampling at points makes the amount of data manageable. The measurements are then converted

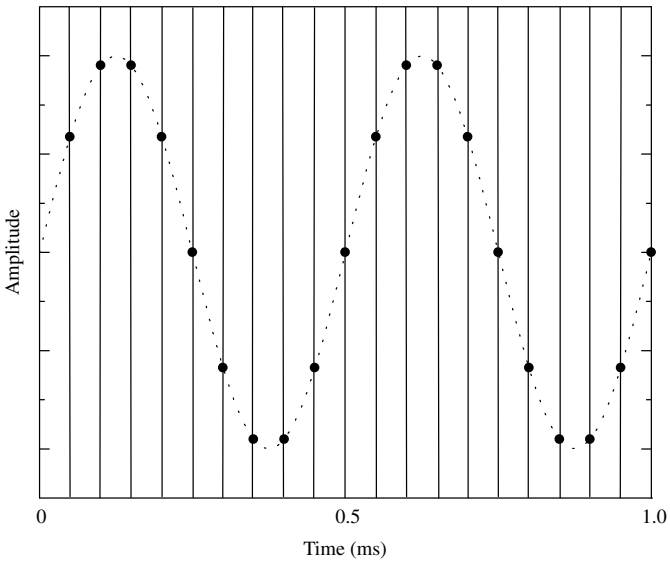


Figure 2.8 Sampling of a waveform. The waveform is measured at points represented by each vertical line. Regions of the waveform in between measurement points are not sampled – all the computer can see are the measured points – so a small amount of error, called quantization error, is introduced by the sampling procedure.

to base two (digital) numbers that a computer can process. The digitizing device has a limit to how many possible amplitude values it can read. The limit depends on the number of digital bits devoted to it, which is usually 16: hence, the number of possible amplitude values is 2^{16}, or 65,536, and this is called 16-bit resolution. Once the computer processes the numbers, the digitized signal can be copied exactly so that the signal doesn't degrade any more.

The next thing to figure out is how often to take the samples. This is called the *sampling rate*. The sampling rate becomes an issue because, if you don't have enough samples, the computer will misinterpret the signal and tell you that there's a frequency that isn't actually there. This misinterpretation is called *aliasing*. Figure 2.9 shows how aliasing happens. If the sampling rate's too low, you'll get samples where the dots are on the figure. The computer just connects the dots and extrapolates a waveform that fits them, which, in this example, didn't exist in the original signal. To prevent aliasing, your sampling rate has to be at least twice the frequency of the highest-frequency component of the signal. Or, to put it another way, the highest-frequency component of the signal can be no higher than half the sampling rate. The frequency that is half the sampling rate is called the *Nyquist frequency*. For recording music, a sampling rate of 44.1 kilohertz (kHz) is most commonly used, so that the Nyquist frequency is 22.05 kHz. However, for recording voices, a sampling rate of 22.05 kHz is adequate because voices don't contain any perceptually important components over 11 kHz.