Making Everything Easier!™

Phonetics

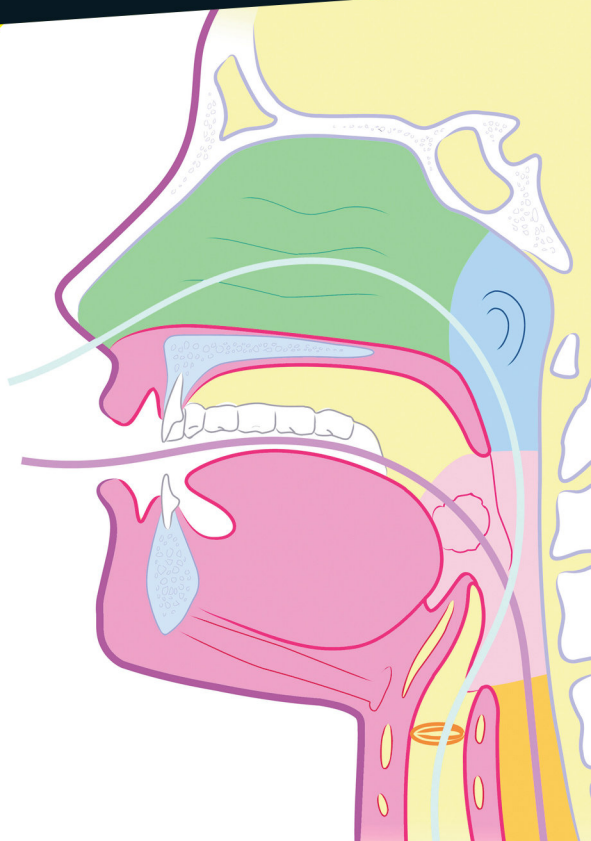
FOR DUMMIES®

A Wiley Brand

Learn to:

- Understand the science behind speech
- Grasp real-world applications of phonetics
- Transcribe speech using the International Phonetic Alphabet (IPA)
- Perform well in your phonetics course

fənɛrɪks



William F. Katz, PhD
Professor, The University of Texas at Dallas

Get More and Do More at Dummies.com®

Start with **FREE** Cheat Sheets

Cheat Sheets include

- Checklists
- Charts
- Common Instructions
- And Other Good Stuff!

To access the Cheat Sheet created specifically for this book, go to
Online Cheat Sheet URL: www.dummies.com/cheatsheet/phonetics

Get Smart at Dummies.com

Dummies.com makes your life easier with 1,000s of answers on everything from removing wallpaper to using the latest version of Windows.

Check out our

- Videos
- Illustrated Articles
- Step-by-Step Instructions

Plus, each month you can win valuable prizes by entering our Dummies.com sweepstakes. *

Want a weekly dose of Dummies? Sign up for Newsletters on

- Digital Photography
- Microsoft Windows & Office
- Personal Finance & Investing
- Health & Wellness
- Computing, iPods & Cell Phones
- eBay
- Internet
- Food, Home & Garden

Find out "HOW" at Dummies.com

**Sweepstakes not currently available in all countries; visit Dummies.com for official rules.*

by William F. Katz, PhD
Professor, The University of Texas at Dallas

FOR DUMMIES
A Wiley Brand

Phonetics For Dummies®

Published by: **John Wiley & Sons, Inc.,** 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

ISBN 978-1-118-50508-3 (pbk); ISBN 978-1-118-50509-0 (ebk); 978-1-118-50510-6 (ebk); 978-1-118-50511-3 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents at a Glance

Table of Contents

Introduction

Welcome to the world of phonetics — the few, the bold, the chosen. You're about to embark on a journey that will enable you to make sounds you never thought possible and to scribble characters in a secret language so that only fellow phoneticians can understand what you're doing. This code, the International Phonetic Alphabet (IPA), is a standard among phoneticians, linguists, teachers, and clinicians worldwide.

Phonetics is the scientific study of the sounds of language. Phonetics includes how speech sounds are produced *(articulatory phonetics),* the physical nature of the sounds themselves *(acoustic phonetics),* and how speech is heard by listeners *(perceptual/linguistic phonetics).*

The information you can gain in an introductory college course on phonetics is essential if you're interested in language learning or teaching. Understanding *phonetic transcription* (that special code language) is critical to anyone pursuing a career in speech language pathology or audiology.

Others can also benefit from studying phonetics. Actors and actresses can greatly improve the convincingness of the characters they portray by adding a basic knowledge of phonetic principles to their background and training. Doing so can make a portrayed accent much more consistent and believable. And if you're a secret drama queen, you can enjoy the fun of trying very different language sounds by using principles of articulatory and acoustic phonetics. No matter what your final career, a basic phonetics class will help you understand how spoken languages work, letting you see the world of speech and language in a whole new light.

About This Book

Phonetics For Dummies gives you an introduction to the scientific study of speech sounds, which includes material from articulatory, acoustic, and perceptual phonetics.

I introduce the field of *phonology* (systems of sound rules in language) and explain how to classify speech sounds using the IPA. I provide examples from foreign accents, dialectology, communication disorders, and children's speech.

I present all the material in a modular format, just like all the other *For Dummies* books, which means you can flip to any chapter or section and read just what you need without having to read anything else. You just need to adhere to some basic ground rules when reading this book and studying phonetics in your class. Here are the big three:

- **Study the facts and theory.** Phonetics covers a broad range of topics, including physiology, acoustics, and perception, which means you need to familiarize yourself with a lot of new terminology. The more you study, the better you'll become.
- **Practice speaking and listening.** An equally important part of being successful is ear training and oral practice (like learning to speak a second language). To get really good at the practical part of the trade, focus on the speaking and listening exercises that I provide throughout the book.
- **Stay persistent and don't give up.** Some principles of phonetics are dead easy, whereas others are trickier. Also, many language sounds can be mastered on the first try, whereas others can even take expert phoneticians (such as Peter Ladefoged) up to 20 years to achieve. Keep at it and the payoff will be worth it!

You can only pack so much into a book nowadays, so I have also recommended many Internet websites that contain more information. These links can be especially helpful for phonetics because multimedia (sound and video) is a powerful tool for mastering speech.

Conventions Used in This Book

This book uses several symbols commonly employed by phoneticians worldwide. If they're new to you, don't worry. They were foreign to even the most expert phoneticians once. Check out these conventions to help you navigate your way through this book (and also in your application of phonetics):

- **/ /:** Angle brackets (or slash marks) denote broad, *phonemic* (indicating only sounds that are meaningful in a language) transcription.
- **[ ]:** Square brackets mark narrow, phonetic transcription. This more detailed representation captures language-particular rules that are part of a language's phonology.

- /kæt/ **or "cat":** This transcription is the International Phonetic Alphabet (IPA) in action. The IPA is a system of notation designed to represent the sounds of the spoken languages of the world. I use the IPA in slash marks (*broad transcription*) for more general description of language sounds (/kæt/), and the IPA in square brackets *(narrow transcription)* to capture greater detail ([k^{h}æt]). I use quotation marks for spelled examples so you don't mistake the letters for IPA symbols.

I use these additional conventions throughout this book. Some are consistent with other *For Dummies* books:

- All Web addresses appear in `monofont`. If you've reading an ebook version, the URLs are live links.
- Some academics seem to feel superior if they use big words that would leave a normal person with a throbbing headache. For example, *anticipatory labial coarticulation* or *intra-oral articulatory undershoot*. Maybe academics just don't get enough love as young children? At any rate, this shouldn't be *your* problem! To spare you the worst of this verbiage, I use *italics* when I clearly define many terms to help you decipher concepts. I also use italics to emphasize stressed syllables or sounds in words, such as "*b*ig" or "*pill*ow".

  I use quotation marks around words that I discuss in different situations, such as when I transcribe them or when I consider sounds. For example, "pillow" /ˈpɪlo/.
- **Bold** is used to highlight the action parts of numbered steps and to emphasize keywords.

Foolish Assumptions

When writing this book, I assume that you're like many of the phonetic students I've worked with for the past 20 years, and share the following traits:

- You're fascinated by language.
- You look forward to discovering more about the speech sounds of the world, but perhaps you have a feeling of chilling dread upon hearing the word *phonetics*.
- You want to be able to describe speech for professional reasons.
- You enjoy hearing different versions of English and telling an Aussie from a Kiwi.
- You're taking an entry-level phonetics class and are completely new to the subject.

If so, then this book is for you. More than likely, you want an introduction to the world of phonetics in an easily accessible fashion that gives you just what you need to know.

What You're Not to Read

Like all *For Dummies* books, this one is organized so that you can find the information that matters to you and ignore the stuff you don't care about. You don't even have to read the chapters in any particular order; each chapter contains the information you need for that chapter's topic, and I provide cross-references if you want to read more about a specific subject. You don't even have to read the entire book — but gosh, don't you want to?

Occasionally, you'll see sidebars, which are shaded boxes of text that go into detail on a particular topic. You don't have to read them unless you're interested; skipping them won't hamper you in understanding the rest of the text. (But I think you'll find them fascinating!)

You can also skip paragraphs marked with the Technical Stuff icon. This information is a tad more technical than what you really need to know to grasp the concept at hand.

How This Book Is Organized

This book is divided into five parts. Here is a rundown of these parts.

Part I: Getting Started with Phonetics

Part I starts with the source-filter model of speech production, describing how individual consonants and vowels are produced. You get to practice, feeling about in your mouth as you do so. I then show how speech sounds are classified using the IPA. This part of the book includes an introduction to phonology, the rules of how speech sounds combine.

Part II: Speculating about English Speech Sounds

Part II shows you further details of English sound production, including processes relevant to narrow transcription. This part focuses on concepts

such as feature theory, phonemes, and allophones — all essential to understanding the relationship between phonetics and phonology. This part also includes information about melody in language, allowing you to analyze languages that sound very different than English and to include prosodic information in your transcriptions.

Part III: Having a Blast: Sound, Waveforms, and Speech Movement

Part III provides grounding in acoustic phonetics, the study of speech sounds themselves. In this part, I begin with sound itself, examining wave theory, sound properties of the vibrating vocal folds, and sound shaping by the lips, jaw, tongue, and velum. I also cover the practical skill of spectrogram reading. You can uncover ways in which speech sounds affect perception (such as voice onset time and formant frequency transitions).

Part IV: Going Global with Phonetics

Part IV branches out with information on languages other than English. These languages have different airstream mechanisms (such as sucking air in to make speech), different states of the voice box (such as making a creaking sound like a toad), and use *phonemic tone* (making high and low sounds to change word meaning). This part also has transcribing examples drawn from children's speech, different varieties of English and productions by individuals with aphasia, dysarthria, and apraxia of speech. The goal is to provide you with a variety of real-world situations for a range of transcribing experiences.

Part V: The Part of Tens

This part seeks to set you straight with some standard lists of ten things. Here I include ten common mistakes that beginning transcribers often make and what you can do to avoid those mishaps. This part also seeks to dispel urban legends circulating among the phonetically non-initiated. You can also find a bonus chapter online at `www.dummies.com/extras/phonetics` for a look at phonetics of the *ph*uture.

Icons Used in This Book

Every *For Dummies* book uses icons, which are small pictures in the margins, to help you enjoy your reading experience. Here are the icons that I use:

When I present helpful information that can make your life a bit easier when studying phonetics, I use this icon.

This icon highlights important pieces of information that I suggest you store away because you'll probably use them on a regular basis.

The study of phonetics is very hands-on. This icon points out different steps and exercises you can do to see (and hear) firsthand phonetics in action. These exercises are fun and show you what your anatomy (your tongue, jaw, lips, and so on) does when making sounds and how you can produce different sounds.

Although everything I write is interesting, not all of it is essential to your understanding the ins and outs of phonetics. If something is nonessential, I use this icon.

This icon alerts you of a potential pitfall or danger.

Where to Go from Here

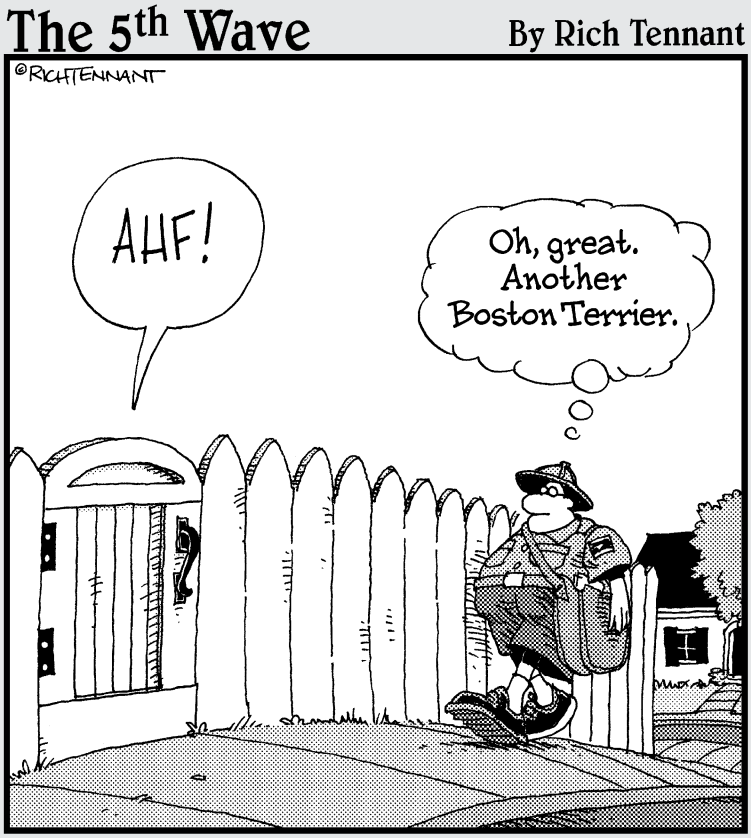
You don't have to read this book in order — feel free to just flip around and focus in on whatever catches your interest. If you're using this book as a way of catching up on a regular college course in phonetics, go to the table of contents or index, search for a topic that interests you, and start reading.

If you'd rather read from the beginning to the end, go for it. Just start with Chapter 1 and start reading. If you want a refresher on the IPA, start with Chapter 3, or if you need to strengthen your knowledge of phonological rules, Chapters 8 and 9 are a good place to begin. No matter where you start, you can find a plethora of valuable information to help with your future phonetic endeavors.

If you want more hands-on practice with your transcriptions, check out some extra multimedia material (located at `www.dummies.com/go/phoneticsfd`) that gives you some exercises and quizzes.

Part I

Getting Started with Phonetics



In this part . . .

- Get the complete lowdown on what phonetics is and why so many different fields study it.
- Familiarize yourself with all the human anatomy that play important role in phonetics, including the lips, tongue, larynx, and vocal folds.
- Understand how the different parts of anatomy work together to produce individual consonants, vowels, syllables, and words.
- Examine the different parts of the *International Phonetic Alphabet (IPA)* to see how phoneticians use it to transcribe spoken speech and begin to make your own transcriptions.
- Identify how different speech sounds are classified and the importance of *voicing* (whether the vocal folds are buzzing), *places of articulation* (the location in your mouth where consonants are formed), and *manner of articulation* (how consonants are formed).
- See how sounds are broken down to the most basic level (phonemes) and how they work together to form words.

Chapter 1

Understanding the A-B-Cs of Phonetics

In This Chapter

- Nurturing your inner phonetician
- Embracing phonetics, not fearing it
- Deciding to prescribe or describe

People talk all day long and never think about it until something goes wrong. For example, a person may suddenly say something completely pointless or embarrassing. A slip of the tongue can cause words or a phrase to come out wrong. Phonetics helps you appreciate many things about how speech is produced and how speech breaks down.

This chapter serves as a jumping-off point into the world of phonetics. Here you can see that phonetics can do the following:

- Provide a systematic means for transcribing speech sounds by using the International Phonetic Alphabet (IPA).
- Explain how healthy speech is produced, which is especially important for understanding the problems of people with neurological disorders, such as stroke, brain tumors, or head injury, who may end up with far more involved speech difficulties.
- Help language learners and teachers, particularly instructors of English as a second language, better understand the sounds of foreign languages so they can be understood.
- Give actors needing to portray different varieties of English (such as American, Australian, British, Caribbean, or New Zealand) the principles of how sounds are produced and how different English accents are characterized.

This chapter serves as a quick overview to your phonetics course. Use it to get your feet wet in phonetics and *phonology,* the way that sounds pattern systematically in language.

Speaking the Truth about Phonetics

> *"The history of phonetics — going back some 2.5 millennia — makes it perhaps the oldest of the behavioral sciences and, given the longevity and applicability of some of the early findings from these times, one of the most successful"*
>
> — Professor John Ohala, University of California, Berkeley

When I tell people that I'm a phonetician, they sometimes respond by saying *a what?* Once in a rare while, they know what phonetics is and tell me how much they enjoyed studying it in college. These people are typically language lovers — folks who enjoy studying foreign tongues, travelling, and experiencing different cultures.

Unfortunately, some people react negatively and share their horror stories of having taken a phonetics course during college. Despite its astounding success among the behavioral sciences, phonetics has received disdain from some students because of these reasons:

- **A lot of specialized jargon and technical terminology:** In phonetics, you need to know some biology, including names for body parts and the physiology of speech. You also need to know some physics, such as the basics of acoustics and speech waveforms. In addition, phonetics involves many social and psychological words, for example when discussing *speech perception* (the study of how language sounds are heard and understood) and *dialectology* (the study of language regional differences). Having to master all this jargon can cause some students to feel that phonetics is hard and quickly become discouraged.
- **Speaking and ear training skills:** When studying phonetics, you must practice speaking and listening to new sounds. For anyone who already experienced second language learning (or enjoys music or singing), doing so isn't a big deal. However, if you're caught off guard by this expectation from the get-go, you may underestimate the amount and type of work involved.
- **The stigma of being a phonetician:** Phoneticians and linguists are often unfairly viewed as nit-picking types who enjoy bossing people around by telling them how to talk. With this kind of role model, working on phonetics can sometimes seems about as exciting as ironing or watching water boil.

I beg to differ with these reasons. Yes, phonetics does have a lot of technical terms, but hang in there and take the time to figure out what they mean because it will be worth your time. With phonetics, consider listening and speaking the different sounds as a fun activity. Working in the field of phonetics is actually an enjoyable and exciting one. Refer to the later section, "Finding Phonetic Solutions to the Problems of the World" and see what impact phonetics has in everyday speech.

Prescribing and Describing: A Modern Balance

This idea that *linguists* (those who study language) and *phoneticians* (those who work with speech sounds) are out to change your language comes from a tradition called *prescriptivism*, which means judging what is correct. Many of the founders of the field of modern phonetics, including Daniel Jones and Henry Sweet, have relied on this tradition. You may be familiar with phoneticians taking this position, for example, the character of Henry Higgins, in the play *Pygmalion* and the musical *My Fair Lady,* or Lionel Logue, as portrayed in the more recent film, *The King's Speech.* At this time and place (England in early 1900s) phoneticians earned their keep mainly by teaching people how to speak "properly."

However, much has changed since then. In general, *linguistics* (the study of language) has broadened to include not only studies close to literature and the humanities (called *philology*, or love of language), but also to disciplines within the cognitive sciences. Thus, linguistics is often taught not only in literature departments, but also in psychology and neural science groups.

These changes have also affected the field of phonetics. Overall, phoneticians have learned to listen more and correct less. Current phonetics is largely *descriptive* (observing how different languages and accents sound), instead of being prescriptive. Descriptive phoneticians are content to identify the factors responsible for spoken language variation (such as social or geographic differences) and to not necessarily translate this knowledge into scolding others as to how they *should* sound.

You can see evidence of this descriptive attitude in the term *General American English (GAE),* used throughout this book, when talking about American norms. (GAE basically means a major accent of American English, most similar to a generalized Midwestern accent; check out Chapter 18 for more information about it.) Although the difference may seem subtle, GAE has a very different flavor than a label such as *Standard American English (SAE),* used by some authors to refer to the same accent. After all, if someone is *standard*, what might that make you or me? Substandard? You can see how the idea of an accent standard carries the sense of prescription, making some folks uneasy.

Scientifically, descriptivism is the way to go. This viewpoint permits phoneticians to study language and speech without the baggage of having to tell people how they should sound. Other spokespeople in society may take a presciptivist position and recommend that certain words, pronunciations, or usages be promoted over others. This prescriptivism is generally based on the idea that language values should be preserved and that nobody wants to speak a language that doesn't have correct forms.

Finding Phonetic Solutions to the Problems of the World

Phonetics can help a lot of problems related to speech. You may be surprised at how omnipresent phonetics is in everyday speech. If you're taking a phonetics course or you're reading to discover more about language and you come across a perplexing problem, the following can refer you to the chapter in this book where I address the solutions.

- How does my body produce speech? Check out Chapter 2.
- I have seen these symbols: /ʒ/, /ʧ/, /ə/, /θ/, /ɚ/, /æ/, /ŋ/, /ʌ/, and /ʊ/. What are they? Refer to Chapter 3.
- Why do Chinese and Vietnamese people sound like their voices are going up and down when they speak? Head to Chapter 3.
- What happens in my throat when I speak, whisper, or sing? Flip to Chapter 4.
- How are speech sounds classified? Check out Chapter 5.
- I have taken a phonetics course, but I *still* don't understand the ideas of *phoneme* and *allophone*. What are they? Refer to Chapter 5.
- What exactly is a glottal stop? Go to Chapter 6.
- What is coarticulation? Does it always occur? Flip to Chapter 6.
- How are vowels produced differently in British and American English? Check out Chapter 7.
- Is it okay to drop my "R"s? Head to Chapter 7.
- What exactly is phonology? Go to Chapter 8.
- Do all people in the world have the same kind of sound changes in their languages? Check out Chapter 8.
- How do I apply diacritics in transcription? Chapter 9 can help.
- I need to know how to narrowly transcribe English. What do I do? Look in Chapter 9.
- How do I transcribe speech that is all run together? Head to Chapter 10.
- What role does melody play in speech? Go to Chapter 10.
- How do I mark speech melody in my transcriptions? Check out Chapter 11.

- How is speech described at the level of sound? Refer to Chapter 12.
- How can I use computer programs to analyze speech? Look in Chapter 12.
- My teacher asked me to decode a sound spectrogram, and I am stuck. What do I do? Chapter 13 can help.
- How do people perceive speech? Refer to Chapter 14.
- Why do speakers of different languages make those odd creaky and breathy sounds? Go to Chapter 15.
- What is voice onset time (VOT)? Chapter 15 has what you need.
- How do speakers of other languages make those peculiar r-like sounds? What about guttural sounds at the backs of their throats and clicks? Look in Chapter 16.
- Are some consonants held longer than others? What about some vowels? Refer to Chapter 16.
- How do I transcribe child language? Check out Chapter 17.
- How can you tell normal child speech from child speech that is delayed or disordered? Go to Chapter 17.
- What exactly are the differences between British, Australian, and New Zealand English? I just opened my mouth and inserted my foot. Chapter 18 can help ease your problems.
- Can you show me some examples of aphasia, apraxia, and dysarthria transcribed? Head to Chapter 19.
- I make mistakes when I transcribe. What can I do to improve? Chapter 20 discusses ten of the most common mistakes that people make when transcribing, and what you can do to avoid them.
- How can I know when someone is telling an urban myth about English accents? Zip to Chapter 21.

Chapter 2

The Lowdown on the Science of Speech Sounds

In This Chapter

- Spelling out what phonetics and phonology are
- Understanding how speech sounds are made
- Recognizing speech anatomy, up close and personal

Phonetics is centrally concerned with speech, a uniquely human behavior. Animals may bark, squeak, or meow to communicate. Parrots and mynah birds can imitate speech and even follow limited sets of human commands. However, only people naturally use speech to communicate. As the philosopher Bertrand Russell put it, "No matter how eloquently a dog may bark, he cannot tell you that his parents were poor, but honest."

In this chapter, I introduce you to the basic way in which speech is produced. I explain the source-filter theory of speech production and the key parts of your anatomy responsible for carrying it out. I begin picking up key features that phoneticians use to describe speech sounds, such as voicing, place of articulation, and manner of articulation.

Phoneticians *transcribe* (write down) speech sounds of any language in the world using special symbols that are part of the International Phonetic Alphabet (IPA). Throughout this book, I walk you through more and more of these IPA symbols, until transcription becomes a cinch. For now, I am careful to indicate spelled words in quotes (such as "bee") and their IPA symbols in slash marks, meaning broad transcription, such as /bi/. (Refer to Chapter 3 for in-depth information on the IPA.)

Defining Phonetics and Phonology

Phonetics is the scientific study of the sounds of language. You may recognize the root *phon-* meaning sound (as in "telephone"). However, phonetics doesn't refer to just any sort of sound (such as a door slamming). Rather, it deals specifically with the sounds of spoken human language. As such, it's part of the larger field of linguistics, the scientific study of language. (Check out *Linguistics For Dummies* by Rose-Marie Dechaine PhD, Strang Burton PhD, and Eric Vatikiotis-Bateson PhD [John Wiley & Sons, Inc.] for more information.)

Phonetics is closely related to *phonology*, the study of the sound systems and rules in language. The difference between phonetics and phonology can seem a bit tricky at first, but it's actually pretty straightforward. Phonetics deals with the sounds themselves. The more complicated part is the rules and systems (phonology). All languages have sound rules. They're not explicit (such as "Keep off the grass!"), but instead they're implicit or effortlessly understood.

To get a basic idea of phonological rules, try a simple exercise. Fill in the opposite of these three English words. (I did the first one for you.)

tolerant	*in*tolerant
consistent	_____
possible	_____

You probably answered "*in*consistent" and "*im*possible," right? Here's the issue. The prefix "in" means "not" (or opposite) in English, so why does the "in" change to "im" for "*im*possible?" It does so because of a sound rule. In this case, the phonological rule is known as *assimilation* (one sound becoming more like another). In this example, a key consonant changes from one made with the tongue (the "n" sound) to one made at the lips (the "m" sound) in order to match the "p" sound of "possible," also produced at the lips. The effect of this phonological rule is to make speech easier to produce. To get a feel for this, try to say "*in*-possible" three times rapidly in succession. Now, try "*im*possible." You can see that saying "impossible" is easier.

I focus more discussion on phonology in Chapters 8 and 9. Now you just need to know that phonological rules are an important part of all spoken languages. One of the key goals of phonology is to figure out which rules are language-specific (applying only to that language) and which are universal.

Phoneticians specialize in describing and understanding speech sounds. A phonetician typically has a good ear for hearing languages and accents, is skilled in the use of computer programs for speech analysis, can analyze speech movement or physiology, and can transcribe using the IPA.

Because phonetics and phonology are closely allied disciplines, a phonetician typically knows some phonology, and a phonologist is grounded in phonetics, even though their main objects of study are somewhat different.

Phonetics can tell people about what language sounds are, how language sounds are produced, and how to transcribe these sounds for many purposes. Phonetics is important for a wide variety of fields, including computer speech and language processing, speech and language pathology, language instruction, acting, voice-over coaching, dialectology, and forensics.

A big part of a person's identity is how you sound when you speak — phonetics lets you understand this in a whole new way. And it's true what the experts say: Phonetics is definitely helpful for anyone learning a new language.

Sourcing and Filtering: How People Make Speech

Scientists have long wondered exactly how speech is produced. Our current best explanation is called the *source-filter theory,* also known as the *acoustic theory of speech production*. The source-filter theory best explains how speech works.

The idea behind this theory is that speech begins with a breathy exhalation from the lungs, causing raw sound to be generated in the throat. This sound-generating activity is the *source.* The source may consist of buzzing of the *vocal folds* (also known as the *vocal cords*), which sounds like an ordinary voice. The source may also include hissing noise, which sounds like a whisper. The movement of the lips, tongue, and jaw (for oral sounds) and the use of the nose (for nasal sounds) shapes this raw sound and is the part of the system known as the *filter.*

The raw sound is filtered into something recognizable. A filter is anything that can selectively permit some things to pass through and block other things (kind of like what your coffee filter does). In this case, the filter allows some frequencies of sound to pass through, while blocking others.

After raw sound is created by a buzzing larynx and/or hissing noise, the sound is filtered by passing through differently shaped airway channels formed by the movement of the speech *articulators* (tongue, lips, jaw, and velum). This sound-shaping process results in fully formed speech (see Figure 2-1 for what this looks like).

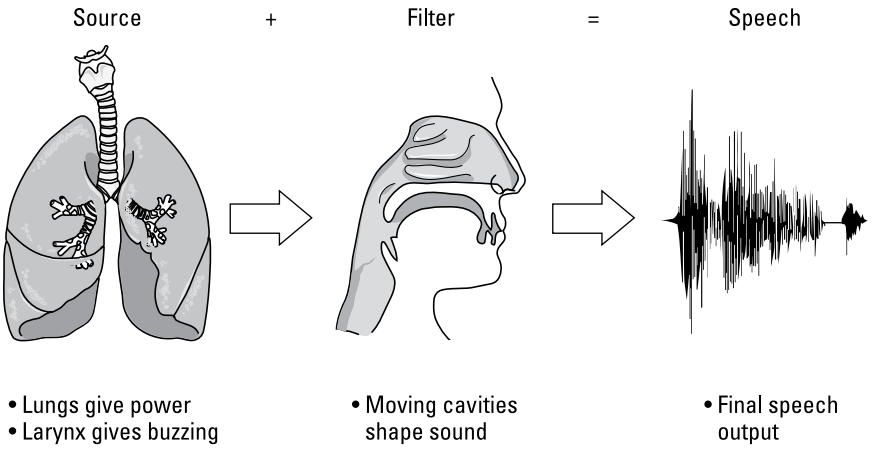


Figure 2-1: The source-filter theory of speech production in action.

Illustration by Wiley, Composition Services Graphics

Let me give you an analogy to help you understand. The first part of the speaking process is like the mouthpiece of a wind instrument, converting air pressure into sound. The filter is the main part of a wind instrument; no one simply plays a mouthpiece. Some kind of instrument body (such as a saxophone or flute) must form the musical sound. Similarly, you start talking with a vibrating source (your vocal folds). You then shape the sound with the instrument of your moving articulators, as the filter.

Here are a few other important points to remember with the source-filter theory:

- The source and filter are largely independent of each other. A talker can have problems with one part of the system, while the other part remains intact.
- The voicing source can be affected by laryngitis (as in a common cold), more serious disease (such as cancers), injuries, or paralysis.
- An alternative voicing source, such as an external artificial larynx, can provide voicing if the vocal folds are no longer able to function.
- The sources and filters of men and women differ. Overall, men have lower voices (different source characteristics) and different filter shapes (created by the mouth and throat passageways) than women.

Thankfully, people never have to really think about making these shapes. If so, imagine how people would ever be able to talk. Nevertheless, this theory explains how humans do talk. It's quite different than, say, rubbing a raspy limb across your body (like the katydid) or drumming your feet on the ground (like the prairie vole cricket) to communicate.

Gunnar Fant and the source-filter theory

The source-filter theory of speech production was the brainchild of Gunnar Fant (1919–2009), a pioneering Swedish professor of speech communication. After earning a master's degree in engineering at Stockholm at the end of World War II, Fant began to apply this knowledge to analyze and synthesize speech sounds. His doctoral dissertation, the *Acoustic Theory of Speech Production*, soon became an international standard. Fant's research led to the development of whole new technologies, including computer speech synthesis, and helped make phonetics more available to a variety of professions. At age 81, while still working actively on phonetics research, Fant wrote in "Half a Century in Phonetics and Speech Research,":

> "Mankind is making much progress in mapping the genetic code. We need some of the same patience and persistence in mapping the speech code."

Getting Acquainted with Your Speaking System

Although most people speak all their lives without really thinking about how they do it, phonetics begins with a close analysis of the speaking system. This part of phonetics, called *articulatory phonetics,* deals with the movement and physiology of speech. However, don't fear — you don't need to be a master phonetician to get this part of the field. In fact, the best way is to pay close attention to your own tongue, lips, jaw, and velum when you speak. As you get better acquainted with your speaking system, the basics of articulatory phonetics should become clear.

Figure 2-2 shows the broad divisions of the speaking system. Researchers divide the system into three levels, separated at the larynx. The lungs, responsible for the breathy source, are below the larynx. The next division is the larynx itself. Buzzing at this part of the body causes voiced sounds, such as in the vowel "ah"' of "h*o*t" (written in IPA characters as /ɑ/) or the sound /z/ of "*z*ip." Finally, the parts of the body that shape sound (the tongue, lips, jaw, and velum) are located above the larynx and are therefore called *supralaryngeal*.

In the following sections, I delve deeper into the different parts of the speech production system and what those parts do to help in the creation of sound. I also walk you through some exercises so you can see by doing — feeling the motion of the lungs, vocal folds, tongue, lips, jaw, and velum, through speech examples.

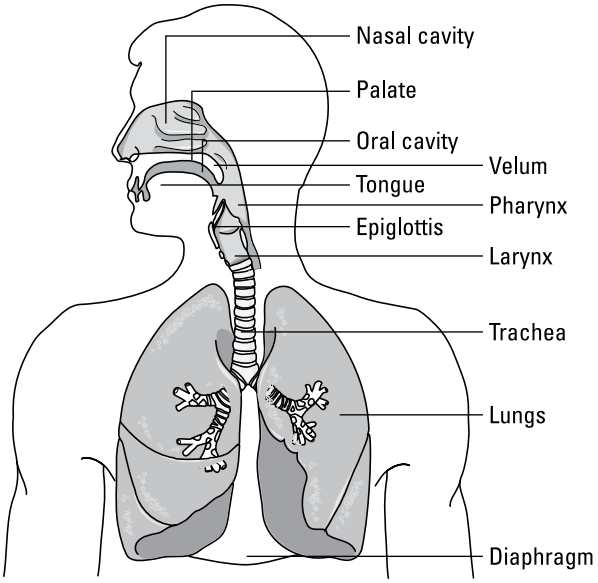


Illustration by Wiley, Composition Services Graphics

Figure 2-2: The main components of the speech production system.

If you're a shy person, you may want to close the door, because some of these exercises can sound, well, embarrassing. On the other hand, if you're a more outgoing type, you can probably enjoy this opportunity to release your inner phonetician.

Powering up your lungs

Speech begins with your lungs. For anyone who has been asked to speak just after an exhausting physical event (say, a marathon), it should come as no surprise that it can be difficult to get words out.

Lung power is important in terms of studying speech sounds for several reasons: Individuals with weakened lungs have characteristic speech difficulties, which is an important part of the study of speech language pathology. Furthermore, as I discuss in Chapter 10, an important feature of speech called *stress* is controlled in large part by how loud a sound is — this, in turn, relates to how much air is puffed out by the lungs.

The role of the lungs in breathing and speech

Your lungs clearly aren't designed to serve only speech. They're part of the respiratory system, designed to bring in oxygen and remove carbon dioxide. Breathing typically begins with the nose, where air is filtered, warmed, and

moistened. Air then moves to the *pharynx*, the part of the throat just behind the nose and into the *trachea*, the so-called *windpipe* that lies in front of the esophagus (or the *food tube*). From the trachea, the tubes split into two bronchi (left and right), then into many *bronchioles* (tiny bronchi), and finally ending up in tiny air sacs called *alveoli*. The gas exchange takes place in these sacs.

When you breathe for speaking, you go into a special mode that is very different than when you walk, run, or just sit around. Basically, speech breathing involves taking in a big breath, then holding back or checking the exhalation process so that enough pressure allows for buzzing at the larynx (also known as *voicing*). If you don't have a steady flow of pressure at the level of the larynx, you can't produce the voiced sounds, which include all the vowels and half of the consonants.

Young children take time to get the timing of this speech breathing right; think of how often you may have heard young kids say overly short breath-group phrases, such as this example:

> "so like Joey got a . . . got a candy and a . . . nice picture from his uncle"

Here the child talker quite literally runs out of breath before finishing his thought.

Some interesting bits about the lungs can give you some more insight into these powerhouse organs:

- They're light and spongy, and they can float on water.
- They contain about two liters (three quarts) of air, fully inflated.
- Your left and right lungs aren't exactly the same. The left lung is divided into two lobes, and the lung on your right side is divided into three. The left lung is also slightly smaller, allowing room for your heart.

When resting, the average adult breathes around 12 to 20 times a minute, which adds up to a total of about 11,000 liters (or 11,623 quarts) of air every day.

Testing your own lung power

You can test your lung power by producing a sustained vowel. To test your lung power, sit up, take a deep breath, and produce the vowel /ɑ/, as in the word "hot," holding it as long as you can. The vowel /ɑ/ is part of the IPA, which I discuss in Chapter 3.

How did you do? Most healthy men can sustain a vowel for around 25 to 35 seconds, and women for 15 to 25 seconds. Next, try the same vowel exercise while lying flat on your back (called being *supine*). You probably can't go on

as long as you did when you were sitting up, and the task should be harder. Due to gravity and biomechanics, the lungs are simply more efficient in certain positions than others. The effect of body position on speech breathing is important to many medical fields, such as speech language pathology.

Buzzing with the vocal folds in the larynx

The *larynx,* a cartilaginous structure sometimes called the *voice box,* is the part of the body responsible for making all voiced sounds. The larynx is a series of cartilages held together by various ligaments and membranes, and also interwoven by a series of muscles. The most important muscles are the vocal folds, two muscular flaps that control the miraculous process of voicing.

Figure 2-3 shows a midsection image of the head. In this figure, you can see the positions of the nasal cavity, oral cavity, pharynx, and larynx. Look to see where the vocal folds and glottis are located. The vocal folds (also known as the *vocal cords*) are located in the larynx. You can find the larynx in the figure at the upper part of the air passage.

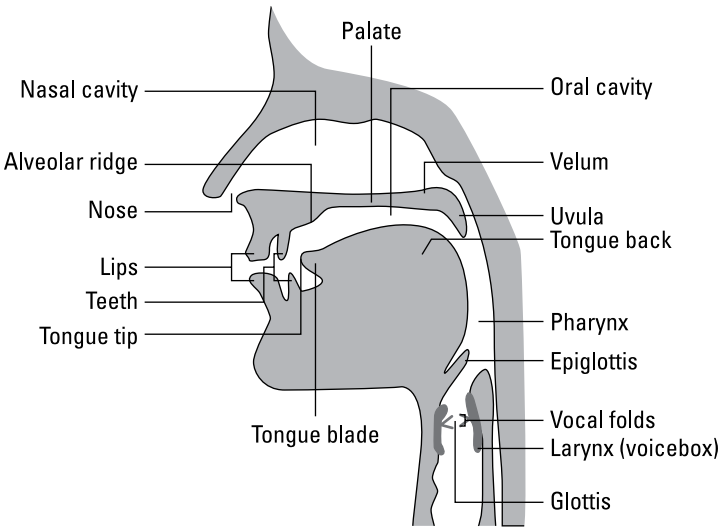


Illustration by Wiley, Composition Services Graphics

Figure 2-3: The midsagittal view of the vocal tract.

The following sections provide some examples you can do to help you get better acquainted with your larynx and glottis.

Getting a buzz from a different source

A common surgery used for the treatment of laryngeal cancer is *laryngectomy,* which is the complete or partial removal of the larynx and vocal folds. After such a surgery, several methods can be used to help a patient speak. One way is to train patients to use an *electrolarynx,* a mechanical (buzzing) device held against the throat to provide vibrations for speech. For laryngectomy patients, the electrolarynx has the advantage of being simple and accessible pre- and post-operation. However, a disadvantage is the rather mechanical voice that results (see `www.youtube.com/watch?v=v55NAjqltEI`).

For phonetics students, trying out an electrolarynx is a fun way to really get the idea of the independence of source and filter. See if you can borrow one from a nearby communication disorders group or clinic. To see how it works, follow these steps:

1. **Place the vibrating membrane against the side of your Adam's apple (laryngeal prominence).**
2. **Turn on the device and silently count to 10.**

   If you did it correctly, you'll feel a pleasant buzzing on your neck while the device *voices* (*phonates*) for you. You may need to try several times to get the coupling just right, so that others can hear you.

Locating your larynx

You can easily find your own larynx. Lightly place your thumb and forefinger on the front of your throat and hold out a vowel. You should feel a buzzing. If you have correctly done it, you're pressing down over the *thyroid cartilage* (refer to the larynx area shown in Figure 2-3) to sense the vibration of the vocal folds while you phonate. If you're male, finding your vocal folds is even more obvious because of your *Adam's apple* (more technically called the *laryngeal prominence*), which is more pronounced in men than women.

Are you happy with your buzzing? Now try saying something else, but this time, whisper. When whispering, switch from a voiced *(phonated)* sound to voiceless. Doing these exercises gives you a good idea of voicing, which is the first of three key features that phoneticians use to classify the speech sounds of the world. (Refer to Chapter 5 for these three key features.) Voicing is one of the most straightforward features for beginning phonetics students because you can always place your hand up to the throat to determine whether a sound is being produced with a voiced source or not.

Stopping with your glottis

Meanwhile, the *glottis* is the empty space between the two vocal folds when they're held open for breathing or for speech. That is, it's basically an empty hole. Your glottis is probably the most important open space in your body

because it regulates air coming in and out of the lungs. Even if you're otherwise able to breathe just fine, if your glottis is clamped shut, air can't enter the lungs.

Clamping your glottis shut is a dangerous situation, so don't try it for long. Nevertheless, it's fun and instructive to try something called the *glottal stop,* /ʔ/, a temporary closing (also called an *adduction*) of the vocal folds that occurs commonly during speech. Are you ready? Stick to these steps as you try this exercise:

1. **Say "uh-oh," loudly and slowly several times.**

   Young children like saying this expression as they are about to drop something expensive (say, your new cell phone) on a cement floor.

2. **Feel your vocal folds clamp shut at the end of "uh," and then open again (the technical term is *abduct*) when you begin saying "oh."**

3. **Try holding the closing gesture (the adduction) after the "uh."**

   You should soon begin feeling uncomfortable and *anoxic* (which means without oxygen) because no air can get to your lungs.

4. **Breathe again, please!**

   I need you alive and healthy to complete these exercises.

5. **Practice by saying other sounds, such as "oh-oh," "ah-ah," and "eeh-eeh," each time holding the glottal stop (at will) across the different vowels.**

   This skill comes in handy when I discuss more about glottal stops used in American English and in different English dialects worldwide in Chapter 18.

Shaping the airflow

Parts of the body filter sound by creating airway shapes above the larynx. Air flowing through differently shaped vessels produces changing speech sounds. Imagine blowing into variously shaped bottles; they don't all sound the same, right? Or consider all the different sizes and shapes of instruments in an orchestra; different shapes lead to different sounds. For this reason, it's important to understand how the movement of your body can shape the air passages in your throat, mouth, and nasal passages in order to produce understandable speech.

Air passages are shaped by the *speech organs*, also known as *articulators*. Phoneticians classify articulators as *movable* (such as the tongue, lips, jaw, and velum) and *fixed* (such as the teeth, alveolar ridge, and hard palate), according to their role in producing sound. Refer to Figures 2-2 and 2-3 to see where the articulators are located.

The movable articulators are as follows. Here you can find some helpful information to understand how each one works:

- **Tongue:** The tongue is the most important articulator, similar in structure to an elephant's trunk. The tongue is a *muscular hydrostat,* which means it's a muscle with a constant volume. (This characteristic is important in the science of making sound because muscular hydrostats are physiologically complex, requiring muscles to work *antagonistically*, against each other, in order to stretch or bend. Such complexity appears necessary for the motor tasks of speech.) The tongue elongates when it extends and bunches up when it contracts. You never directly see the main part of the tongue (the body and root). You can only view the thinner sections (tip/blade/dorsum) when it's extended for viewing. However, scientists can use imaging technologies such as ultrasound, videoflouroscopy, and magnetic resonance imaging to know what these tongue parts look like and how they behave.
- **Jaw:** Although classified as a movable speech articulator, the jaw isn't as important as the tongue. The jaw basically serves as a platform to position the tongue.
- **Lips:** The lips are used mostly to lower vowel sounds through extension. The *lip extension* is also known as *protrusion* or *rounding.* The lips protrude approximately a quarter inch when rounded. English has two rounded vowels, /u/ (as in "b*oo*t"), and /ʊ/ (as in "b*oo*k"). Other languages have more rounded sounds, such as Swedish, French, and German (refer to Chapter 15). These languages require more precise lip rounding than English.

  Lips can also flare and spread (widen). This acts like the bell of a brass instrument to brighten up certain sounds (like /i/ in "b*ea*d").
- **Velum:** The *velum,* also known as the *soft palate,* is fleshy, moveable, and made of muscle. The velum regulates the nasality of speech sounds (for example, /d/ versus /n/, as in the words "*d*ice" and "*n*ice"). The velum makes up the rear third of the roof of the mouth and ends with a hanging body called the *uvula,* which means "bundle of grapes," just in front of the throat.

Some parts of the body are more passive or static during sound production. These so-called fixed articulators are as follows:

- **Teeth:** Your teeth are used to produce the "th" sounds in English, including the voiced consonant /ð/ (as in "*th*ose") and the voiceless consonant /θ/ (as in "*th*ick"). The consonants made here are called *dental.* Your teeth are helpful in making *fricatives,* hissy sounds in which air is forced through a narrow groove, especially /s/, /z/, /f/, and /v/ — like in the words "*s*o," "*z*ip", "*f*eel," and "*v*ote". Tooth loss can affect other speech sounds, including the affricates /tʃ/ (as in "*ch*op") and /dʒ/ (as in "*J*oe").
- **Alveolar ridge:** This is a pronounced body ridge located about a quarter of an inch behind your top teeth. Consonants made here are called *alveolar.*

  You can easily feel the alveolar ridge with your tongue. Say "na-na" or "da-da," and feel where your tongue touches on the roof of your mouth.

  The alveolar ridge is particularly important for producing consonants, including /t/, /d/, /s/, /z/, /n/, /l/, and /ɹ/, as in the words "*t*ime," "*d*ime," "*s*ick," "*z*oo," "*n*ice, "*l*ice," and "*r*ice." Many scientists think an exaggerated alveolar ridge has evolved in modern humans to support speech.
- **Hard palate:** It continues just behind the alveolar ridge and makes up the first two-thirds of the roof of your mouth. It's fixed and immovable because it's backed by bone. Consonants made here are called *palatal.* The English consonant /j/ (as in "*y*ellow") is produced at the hard palate.

Producing Consonants

A *consonant* is a sound made by partially or totally blocking the vocal tract during speech production. Consonants are classified based on *where* they're made in the articulatory system (place of articulation), *how* they are produced (manner of articulation), and whether they're *voiced* (made with buzzing of the larynx) or not. These sections discuss the different ways English consonants are made. Remember, each language has its own set of consonants. So English, for example, doesn't have the "rolled r" found in Spanish, and Spanish doesn't have the consonant /dʒ/ as in "ju*dg*e".

Getting to the right place

Basically consonant sounds use different parts of the tongue and the lips. Figure 2-4 shows a midsagittal view of the head, including the lips, tongue, and the consonantal places of articulation.

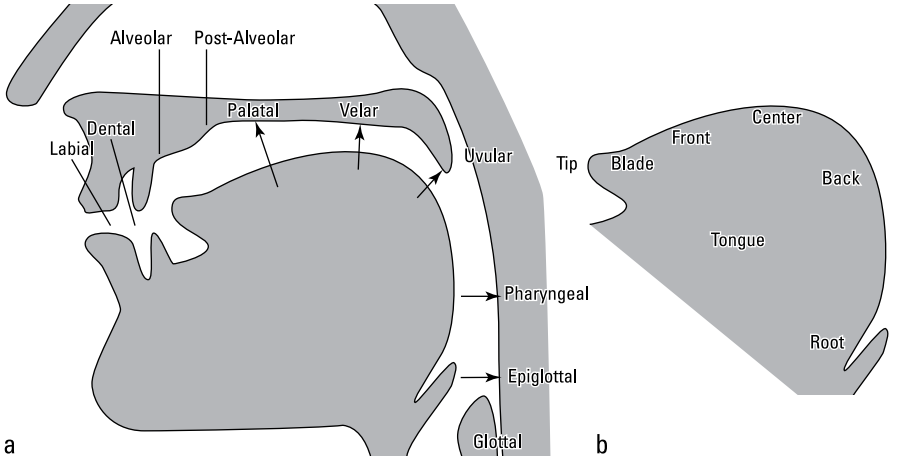


Figure 2-4: The consonantal places of articulation (a) and divisions of the tongue (b).

Illustration by Wiley, Composition Services Graphics

Notice that these regions are relative; there is clearly no "dotted line" separating the front from the back or marking off the tip from the blade (unless you happen to have a disturbing tattoo there, which I doubt). However, these regions play different functional roles in speech. The tip and blade are the most flexible tongue regions. The different parts of the tongue control the sound in the following ways:

- **Coronal:** Speech sounds made using either the tip or blade are called *coronal* (crown-like) sounds.
- **Dorsal:** Speech sounds made using the rear of the tongue are called *dorsal* (back) articulations.

To get an overall feel of what happens when you speak with your lips, tongue, and jaw, slowly say the word "*batik,*" paying attention to where your articulators are as you do so. At the beginning of the word you should sense the separation of the lips for the /b/ (a labial gesture), then the lowering of the tongue and jaw as you pronounce the first syllable. Next, the front of your tongue will rise to make (coronal) contact for the /t/ of "*t*ik." When you reach the end of "ti*k*," you should be able to detect the back (dorsum) of your tongue making (velar) contact with the roof of your mouth for the final /k/ sound.

However, phoneticians typically need to know more detail about where sounds are made than just which parts of the tongue are involved. The following list details the English places of articulation for consonants:

- **Bilabial:** Also called *labial,* sounds made with a constriction at the lips are very common in the languages of the world. Say "*p*at," "*b*at," and "*m*at" to get a good feel for these sounds. Because the lips are a visible part of a

person's body, young children usually use these bilabial sounds in some of their first spoken words ("Momma" or "Poppa"). Think of the baby word terms for mother and father in other languages you may know; they probably contain bilabial consonants.

- **Labiodental:** Your top teeth touch your bottom lip to form these sounds. Say "*f*at" and "*v*at" to sample a voiceless and voiced pair produced at the labiodental place. A person could logically flip things around and try to make a consonant by touching the bottom teeth to the top lip. I can't take any legal responsibility for any spluttering behavior from such an ill-advised anatomical experiment.
- **Dental:** A closure produced at the teeth with contact of the tongue tip and/or blade makes these consonants. For American English, this refers to the "th" sounds, as in "*th*ick" and "*th*is." The first sound is voiceless and is transcribed with the IPA symbol /θ/, *theta*. The second is voiced and is transcribed with the IPA symbol /ð/, *ethe*. Beginning phonetics students frequently mix up /θ/ and /ð/, probably due to the dreadful problem of fixating about spelling. Remember to use your ear and the IPA, and you'll be fine.

Phonetics is a discipline where (for once) you really don't have to worry about how to spell. In fact, an overreliance on spelling can trip you up in many ways. When you hear a word and wish to transcribe it, concentrate on the sounds and don't worry about how it's spelled. Instead, go directly to the IPA characters. If you remain hung up on spelling, a good way to break this habit is to transcribe *nonsense words* also known as *nonce words* because you can't possibly know how they're spelled correctly.

- **Alveolar:** As I discuss in the earlier section, "Shaping the airflow," this important bony ridge on your hard palate makes the sounds /t/, /d/, /s/, /z/, /n/, /l/, and /ɹ/. The tongue tip makes some of these sounds, while the tongue blade makes others.
- **Retroflex:** This name literally means *flexed backwards.* Placing the tongue tip to the rear of the alveolar ridge makes these sounds. Although (as I show you in Chapter 16) such sounds are common in the English accents of India and Pakistan, they're less common in American or British English.
- **Palato-alveolar:** This region is also known as the *post-alveolar.* You make these sounds when you place the tongue blade just behind the alveolar ridge. Constriction is made at the palatal region, as in the sound "sh" of "*sh*ip," transcribed with the IPA character /ʃ/, known as "esh." The voiced equivalent, "zh," as in "plea*s*ure" or "lei*s*ure," is transcribed in the IPA as /ʒ/, "long z" or "yogh." English has many /ʃ/ sounds, but far fewer /ʒ/ sounds (especially because many /ʒ/-containing words are of French or Hungarian origin, thank you, Zsa Zsa Gabor).
- **Palatal:** You make this sound by placing the front of the tongue on the hard palate. It's the loneliest place of articulation in English. Although some languages have many consonants produced here, English has only the gliding sound "y" of "*y*es," transcribed incidentally, with /j/. Repeat "you young yappy yodelers" if you really want a palatal workout.

- **Velar:** For these sounds, you're placing the back of your tongue on the soft palate. That's the pliant, yucky part of the back of your mouth with no underlying bone to make it hard, just cartilage. Try saying "kick" and "gag" to get a mouthful of stop consonants made here. You can also make nasal consonants here, such as the sound at the end of the words "si*ng*, sa*ng*, su*ng*" — transcribed with the IPA symbol /ŋ/, "eng" or "long n."

  Note that /ŋ/ isn't the same nasal consonant as the alveolar /n/, such as in "si*n*." Velar nasals have a much more "back of the mouth" sound than alveolars. Also, people speaking English can't start a word with velar nasals — they occur only at the end of syllables. So, if someone says to you "have a gnice /ŋɑɪs/ day!," you should suspect something has gone terribly, terribly wrong.

  Beginning transcribers may sometimes be confused by "ing" words, such as "thing" (/ɪŋ/ in IPA) or "sang" (/sæŋ/ in IPA). A typical question is "where is the "g"? This is a spelling illusion. Although some speakers may possibly be able to produce a "hard g" (made with a full occlusion) for these examples (for example, "sin*g*"), most talkers don't realize a final stop. They simply end with a velar nasal. Try it and see what you do. On the other hand, if you listen carefully to words, such as "singular," "linguistics," or "wrangle," there indeed should probably be a /g/ placed in the IPA transcription because this sound is produced. I provide more help on problem areas for beginning transcribers in Chapter 20.

Nosing around when you need to

Although it may sound disturbing, people actually talk through their noses at times. The oral airway is connected to the nasal passages — you may have unfortunately discovered this connection if you've unluckily burst out laughing at a funny joke while trying to swallow a sip of soda.

Air usually passes from the lungs through the mouth during speech because during most speech the soft palate raises to close off the passage of air through the nose. However, in the case of nasal consonants, the velum lowers roughly at the same time as the consonantal obstruction in the mouth, resulting in air also flowing out through the nose. People do this miraculous process of shunting air from the oral cavity to the nasal cavity (and back again) automatically, thousands of times each day.

Here is a nifty way to detect nasal airflow during speech. Ladies, get your makeup mirrors! Guys, borrow one from a friend. If the mirror is cool to the touch, you're good to go. If not, place it in a refrigerator for an hour or so, and you'll be ready to try a classic phonetician's trick. Hold the mirror directly under your nose and say "dice" three times. Because the beginning of "*d*ice" has an oral consonant, you should observe, well, *nothing* on the mirror. That is, most air escapes through your mouth for this sound. Next, try saying "nice" three times. This time, you should notice some clear fog marks under each

nostril where your outgoing air during nasal release for /n/ made contact with the mirror. You may now try this with other places of (nasal) articulation, such as the words "mime" and "hang."

Minding your manners

Blocking the vocal tract forms consonants. Forming consonants can happen in different ways: by making a complete closure for a short or long time, by letting air escape in different fashions, or by having the articulators approach each other for a while, resulting in vocal tract shapes that modify airflow. The following list includes some of the main manners of articulation in English. I discuss more details on manner of articulator, including examples for other languages, in Chapters 5 and 16.

- **Stop:** When air is completely blocked during speech, this is called a *stop* consonant. English stops include voiceless consonants /p/, /t/, and /k/ and voiced consonants /b/, /d/, and /g/, as in the words "*p*at," "*t*at," and "*c*at" and "*b*all," "*d*oll," and "*g*all." You make these consonants by blocking airflow in different regions of the mouth. *Nasal stops* (sometimes called just *nasals,* for short) also involve blocking air in the oral cavity, but they're coordinated with a lowering of the velum to allow air to escape through the nose.
- **Fricative:** These consonants all involve producing *friction,* or hissing sound, by bringing two articulators very close to each other and blowing air through. When air passes through a narrow groove or slit, a hiss results (think of opening your car window just a crack while driving down the freeway at a high speed). You hiss with your articulators when you make sounds, such as /f/, /v/, /s/, or /ð/ (as in "*f*at," "*v*at," "*s*at," and "*th*at"). Chapter 6 provides more information on English fricatives.
- **Affricate:** This type of consonant may be thought of as a combination of stop and fricative. That is, an *affricate* starts off sharply with a complete blockage of sound and then transitions into a hiss. As such, the symbols for affricates tend to involve double letters, such as the two affricates found in English, the voiceless /tʃ/ for "*ch*ip" or "whi*ch*," and the voiced affricate /dʒ/, as in "we*dg*e" or "*J*eff." Note that some authors tie the affricate symbols together with a tie or bar, such as /t͡ʃ/, /t‿ʃ/, or /t̄ʃ/. I use more recent conventions and don't do so.
- **Approximant:** In these consonants, two articulators approach or *approximate* each other. As a result, the vocal tract briefly assumes an interesting shape that forms sound without creating any hissing or complete blockage. These sounds tend to have a fluid or "wa-wa"-like quality, and include the English consonants /ɹ/, /l/, /j/, and /w/, as in the words "*r*ake," "*l*ake," "*y*ell," and "*w*ell."

  A good way to remember the English approximants is to think of the phrase *"*your whirlies,*"* because it contains them all: /j/, /ɹ/, /w/, and /l/.

Note that the American English "r" is properly transcribed upside down, /ɹ/, in IPA. Many varieties of "r" sounds exist in the world, and the IPA has reserved the "right side up" symbol, /r/, for the rolling (trilled) "r," for instance in Spanish. I go over more information on IPA characters in Chapter 3.

- **Tap:** For this consonant, sometimes called a *flap,* the tongue makes a single hit against the alveolar ridge. It's a brief voiced event, common in the middle of words such as "city" in American English. A tap is transcribed as /ɾ/ in the IPA.

Producing Vowels

Vowels are produced with relatively little obstruction of air in the vocal tract, which is different than consonants. Phoneticians describe the way in which people produce vowels in different terms than for consonants. Because vowels are made by the tongue being held in rather complicated shapes in various positions, phoneticians settle for rather general expressions such as "high, mid, low" and "front, center, back" to describe vowel place of articulation. Thus, a sound made with the tongue held with the main point of constriction toward the top front of the mouth is called a *high-front vowel,* while a vowel produced with the tongue pretty much in the center of the mouth is called a *mid-central vowel*. The positions of the lips (rounded or not) are also important.

As I describe in Chapters 12 and 13, many phoneticians believe a better description of vowels can be given acoustically, such as what a sound spectrograph measures. Nevertheless, the best way to understand how vowels are formed is to produce them, from the front to the back, and from top to the bottom.

To the front

The *front* vowels are produced with the tongue tip just a bit behind your teeth. Start with the sound "ee" as in "h*ee*d," transcribed in the IPA as /i/. Say this sound three times. This is a high-front vowel because you make it at the very front of your mouth with the tongue pulled as high up as possible. Next, try the words "hid," "hayed," "head," and "had" — in this order. You've just made the front vowel series of American English. In IPA symbols, you transcribe these vowels as /ɪ/, /e/, /ɛ/, and /æ/.

As you speak this series, notice your tongue stays at the front of your mouth, but your tongue and jaw drop because the vowels become progressively lower. By the time you get to "had," you're making a low-front vowel.

To the back

You form the back vowels at, where else, the back of your mouth (big surprise!). Start with "boot" to make /u/, a high-back vowel. Next, please say "book" and "boat." You should feel your tongue lowering in the mouth, with the major constriction still being located at the back. Phoneticians transcribe these vowels of American English as /ʊ/ and /o/.

The next two (low-back) sounds are some of the most difficult to tell apart, so don't panic if you can't immediately decipher them. Say "l*aw*" and "f*a*ther." In most dialects of American English, these words contain the vowels "open-o" (/ɔ/) and /ɑ/, respectively. Most students (and even many phoneticians) have difficulty differentiating between them. These vowels also are merging in many English dialects, making consistent examples difficult to list. For example, some American talkers contrast /ɔ/ and /ɑ/ for "caught-cot", although most don't. Nonetheless, with practice you can get better at sorting out these notorious two vowel sounds at the low-back region of the vowel space!

In the middle: Mid-central vowels

A time-honored method of many phonetics teachers is to save teaching the English central vowels for last because the basics of mid-central vowels are easy, but processing all the details can get a bit involved. For now, let me break them into these two classifications.

"Uh" vowels

The "uh" vowels include the symbols /ə/ "schwa" and /ʌ/ "wedge", as in the words "th*e*" and "m*u*d." Don't be surprised if these two vowels (/ə/ and /ʌ/) sound pretty much the same to you (they do to me) — the difference here has to do with linguistic stress — because words with linguistic content such as nouns, verbs, and adjectives (for example, "mud" and "cut") are produced with greater *linguistic stress* (see Chapter 7 for more details). They're produced with a slightly more open quality and are assigned the symbol /ʌ/. Refer to the later section, "Putting Sounds Together (Suprasegmentals)" for more about linguistic stress. In contrast, English *articles,* such as "the" and "a" (as well as weak syllables in polysyllabic words, such as the "re" in "*re*ply") tend to be produced quietly, that is with less stress. This results in a relatively more closed mouth position for the "uh" sounds, transcribed as the vowel /ə/.

I dislike the names "schwa" and "wedge" because these character names don't represent their intended sounds well. Therefore, I suggest you secretly do what my students do and rename them something like "schwuh" and "wudge." Doing so can help you remember that these symbols represent an "uh" quality.

"Er" vowels

English has /ɚ/ ("r-colored schwa") and /ɝ/ (right-hook reversed epsilon) for "er" mid-central vowels. Notice that both of these characters have a small part on the right (a right hook, not to be confused with the prizefighting gesture) that indicates *rhoticization,* also referred to as *r-coloring.* For most North American accents, you can find the vowels /ɚ/ and /ɝ/ in the words "her" and "shirt."

The good news is that similar stress principles apply with the "er" series as the "uh" series. Pronouns such as "her" or endings such as the "er" in "father" typically don't attract stress and thus are written with an r-colored schwa, /ɚ/. On the other hand, you transcribe a verb, such as "hurt" or an adjective such as "first," with the vowel /ɝ/ (right-hook reversed epsilon).

Embarrassing 'phthongs'?

The vowels in the preceding section are called *monophthongs,* literally "single sound" (in Greek). These vowels have only one sound quality. Try saying "the fat cat on the flat mat." The main words here contain a monophthongal vowel called "ash," written in the IPA as /æ/. Notice how /æ/ vowels have one basic quality — they are, if you will, flat.

Next, try saying the famous phrase "How now brown cow?" Pronounce the phrases slowly and notice that each vowel will seem to slide from an "ah" to an "oo" (or in the IPA, from an /a/ to an /ʊ/). For this reason, these words are each said to each contain a *diphthong,* or a vowel containing two qualities. For /aʊ/, English speakers transition from a low-front to a high-back vowel quality. In addition to /aʊ/, English has two other diphthongs, /aɪ/ (as in "wh*i*te" or "s*i*ze") and /ɔɪ/ (as in "b*oy*" or "l*oi*ter").

Are diphthongs really embarrassing? They shouldn't be, unless you produce them in an exaggerated manner (such as in the previous exercise). However, if you feel shy about producing diphthongs, you may wish to think twice about studying a language, such as the Bern dialect of Swiss German, which has diphthongs and even triphthongs aplenty. Yes, you guessed correctly — in a *triphthong,* one would swing through three different vowel qualities within one vowel-like sound. Check it out with the locals the next time you are in Bern (and don't really worry about being embarrassed).

Putting sounds together (suprasegmentals)

Consonants and vowels are called *segmental* units of speech. When people refer to the consonants and vowels of a language, they're dealing with individual (and logically separable) divisions of speech. This part is an important aspect of phonetics, but surely not the only part. To start with, consonants and vowels combine into *syllables*, an absolutely essentially part of language. Without syllables,

you couldn't even speak your own name (and would, I suppose, be left only with your initials). Therefore, you need to consider larger chunks of language, called *suprasegmentals*, or sections larger than the segment.

Suprasegmentals refer to those features that apply to syllables and larger chunks of language, such as the phrase or sentence. They include changes in *stress* (the relative degree of prominence that a syllable has) and *pitch* (how high or low the sound is), which the following sections explain in greater detail.

Emphasizing a syllable: Linguistic stress

When phoneticians refer to stress, they don't mean emotional stress. For English, *linguistic stress* deals with making a syllable louder, longer, and higher in pitch (that is, making it stand out) compared to others. Stress can serve two different functions in language:

- *Lexical* (or word level)
- *Focus* (or contrastive emphasis)

Part of knowing English is realizing when stress is placed on the correct *syll*able (here at the beginning of the word), and not on a wrong syl*lab*le (such as here, in the middle of the word). Words that are *polysyllabic* (containing more than one syllable) have a correct spot for main stress (also called *primary stress*). Therefore, getting the stress right is an important part of our word learning.

In addition, some English word pairs show regular contrast between nouns and verbs with respect to stress placement. Say these words to yourself:

Noun	***Verb***
record	(to) record
(his) conduct	(to) conduct
(the) permit	(to) permit

You can tell that stress falls on the first syllable of the nouns, and the last syllable of the verbs, right? For some English word pairs stress assignment serves a grammatical role, helping indicate which words are nouns and which are verbs.

Stress can also be used to draw attention (focus) to a certain aspect of an utterance, while downplaying others. Repeat these three sentences, stressing the bolded word in each case:

Sonya plays piano.

Sonya *plays* piano.

Sonya plays *piano.*

Does your stressing these italicized words differently change the meaning of any of these sentences? Each sentence contains the same words — thus, logically, they should all mean the same thing, right? As you probably guessed, they don't. When people stress a certain word in a phrase or sentence, they do shift the emphasis or meaning. These three sentences all seem to answer three different questions:

Who plays piano?	(*Sonya* does!)
Does Sonya listen to piano or play piano?	(She *plays!*)
Does Sonya play the bagpipes?	(No, she plays *piano.*)

Using stress allows people to convey very different emphasis even when using the same words. Correctly using stress in this way is quite a challenge for computers, by the way. Think of how computer speech often sounds or how the stress in your speech may be misunderstood by computerized telephone answering systems.

A good way to practice finding the primary stress of a word is to say it while rapping out the rhythm with your knuckles on a table. For instance, try this with "refrigerator." You should get something like:

knock *knock* knock knock knock

That is, the stress falls on the second syllable ("fridge").

Next, try the word "tendency." You should have:

knock knock knock

Here, stress falls on the first (or initial) syllable.

This method seems to work well for most beginning phonetics students. I think the only time students have difficulty with stress assignment is if they overthink it. Remember, it is a sound thing and really quite simple after you get the hang of it.

Changing how low or high the sound is

Pitch is a suprasegmental feature that results from changes in the rate of buzzing of the larynx. The faster the buzzing, the higher pitched the sound; the slower the buzzing, the lower the sound.

Men and women buzz the larynx at generally different rates. If you're an adult male, on average your larynx buzzes about 120 times per second when you speak. Women and children (having higher voices) buzz at typically about twice that rate, around 220 times per second. This difference is due to the fact that men have larger laryngeal cartilages (Adam's apple) and vocal folds.

Phoneticians call the rate of this buzzing *frequency*, the number of times something completes a cycle over time. In this case, it's the number of times that air pulses from the larynx (resulting from the opening and closing of the vocal folds) per second.

Pitch refers to the way in which frequency is heard. When phoneticians talk about pitch, they aren't referring to the physical means of producing a speech sound, but the way in which a listener is able to place that sound as being higher or lower than another. For example, when people listen to music, they can usually tell when one note is higher or lower than another, although they may not know much else about the music (such as what those notes are or what instruments produced them). Detecting this auditory property of high and low is very important in speech and language.

English uses pitch patterns known as *sentence-level intonation,* which means the way in which pitch changes over a phrase- or sentence-length utterance to affect meaning. Try these two sentences, and listen carefully to the melody as you say each one:

- "I am at the supermarket." This type of simple factual statement is usually produced with a *falling intonation contour.* This means the pitch drops over the course of the sentence, with the word "I" being higher than the word "supermarket." Many phoneticians think this basic type of pitch pattern may be universal (found across the world's languages). People blow off air when they exhale for speech, providing less energy for increased pitch by the end of an utterance, compared to the beginning.
- "Are you eating that egg roll?" In this question, you probably noticed your melody going in the opposite direction, that is — from low to high. In English, people usually form this kind of "yes/no question" (a question that can be answered with a yes or no answer) with a rising intonation pattern. Indeed, if you were to restate the factual sentence "I am eating an egg roll" and change your intonation so that the pitch went from low to high, it would turn into a question or expression of astonishment.

These examples show how a simple switch in intonation contour can change the meaning of words from a statement to a question. In Chapter 10, I discuss more about the power of intonation in English speech.

Chapter 3

Meeting the IPA: Your New Secret Code

In This Chapter

- Taking a closer look at the symbols
- Zipping around the chart
- Recognizing the sounds
- Seeing why the IPA is better than spelling

The *International Phonetic Alphabet* (IPA) is a comprehensive symbol set that lets you transcribe the sounds of any language in the world. The *International Phonetic Association*, a group of phoneticians who meet regularly to adjust features and symbols, revises and maintains the IPA, making sure that all world languages are covered. Many IPA symbols come from Latin characters and resemble English (such as, /b/), so you'll probably feel fairly comfortable with them. However, other symbols may seem foreign to you, such as /ʃ/ or /ŋ/. In this chapter, I show you how to write, understand, and pronounce these IPA characters.

Although most alphabets are designed to represent only one language (or a small set of languages), the IPA represents the sounds of any of the languages in the world. An *alphabet* is any set of letters or symbols in which a language is written. When people speak more specifically of the alphabet, they usually refer to today's system of writing (the ABCs) that has been handed down from the ancient Near East. The word "alphabet" comes from the Greek letters *alpha beta*, and from the Hebrew letters *aleph* and *bet*.

The history of the IPA (not a beer or a terrorist organization)

In 1886, a group of language teachers met in Paris to help school children learn to read and to better pronounce foreign languages. The group eventually became known l' Association Phonétique Internationale (or in English, The International Phonetic Association). The group soon made it its goal to create a universal alphabet to describe the sounds of any language in the world. After 125 years of work and numerous revisions, it came up with today's sophisticated version of the IPA.

Eyeballing the Symbols

When you examine the full IPA chart (see Figure 3-1 or check out `www.langsci.ucl.ac.uk/ipa/IPA_chart_%28C%292005.pdf`), you can see a few hundred different symbols. However, please don't panic! You only need a fraction of them to transcribe English. In these sections, I introduce them to you first. Like the Periodic Table you may have studied in chemistry class, you can also master the basic principles of the IPA chart without getting hung up in all the details. After you master the basics, you can later focus on any other symbols you need.

Each IPA symbol represents unique *voicing* (whether the vocal folds are active during sound production), *place of articulation* (where in the vocal tract a sound is made), and *manner of articulation* (how a sound is produced) for consonants. For vowels, each IPA symbol represents *height* (tongue vertical positioning), *advancement* (tongue horizontal positioning), and *lip-rounding* specifications (whether the lips are protruded for sound production). Refer to Chapter 6 for more information on English consonant features, and Chapter 7 for English vowel features.

Latin alphabet symbols

See if you can begin by spotting the Latin alphabet symbols. They're among the group of symbols labeled with a No. 1 in Figure 3-1, called *pulmonic consonants.* The Latin alphabet symbols include these lower-case characters (/p/, /b/, /m/, /f/, /v/, /t/, /d/, /n/, /s/, /z/, /l/, /c/, /j/, /k/, /g/, /x/, /q/, and /h/), and upper-case characters (/B/, /R/, /G/, /L/, and /N/).

The IPA isn't spelling. Although some of the IPA lower-case Latin symbols may match up pretty well with sounds represented by English letters (for instance, IPA /p/ and the letter "p" in "*p*it"), other IPA Latin symbols (/c/, /j/, /x/, /q/, /B/, /R/, /G/, /L/, and /N/) don't. For instance, IPA /q/ has nothing to do with the letter "q" in *qu*ick or *qu*iet. Rather, /q/ is a throat sound not even found in English but present in Arabic and Sephardic Hebrew.

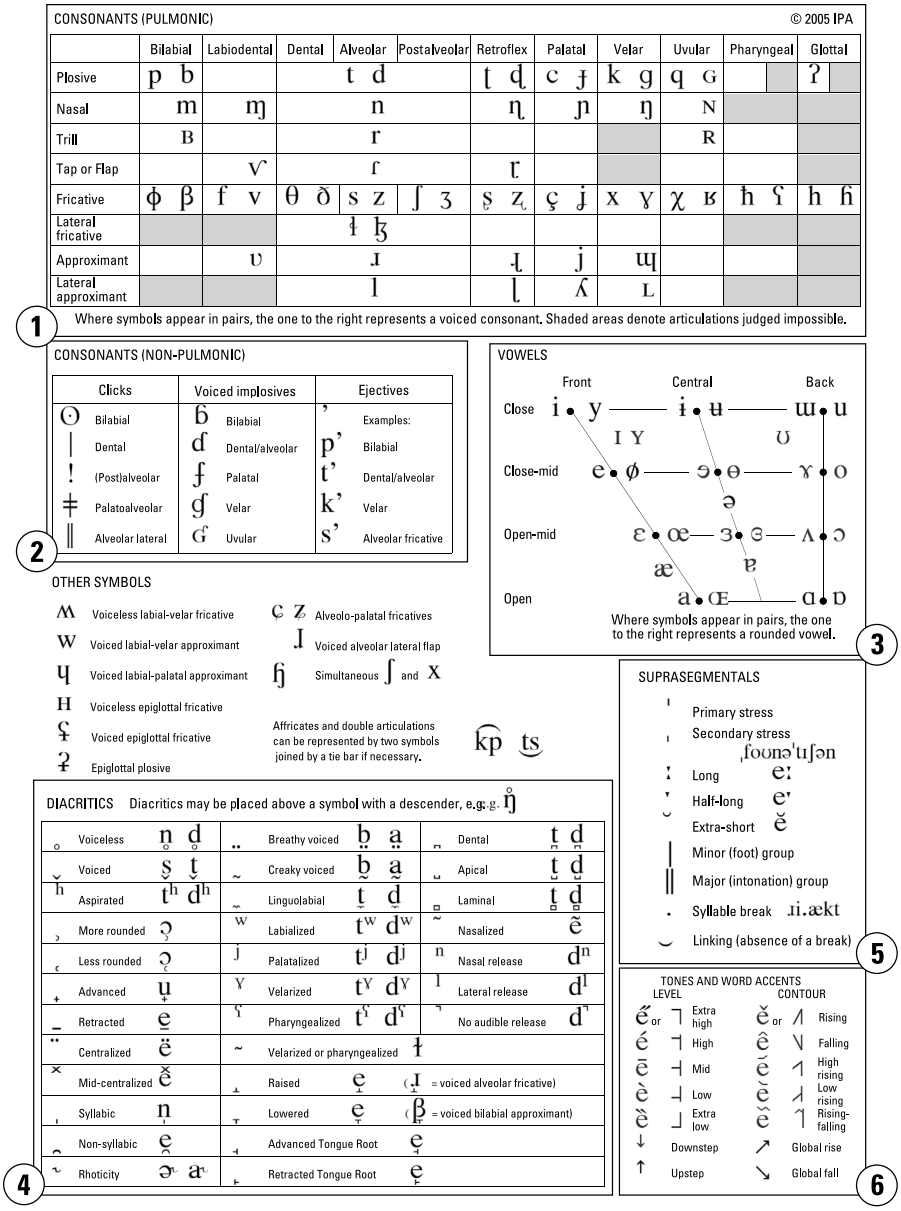


Figure 3-1: The International Phonetic Alphabet (revised to 2005).

You can also find Latin symbols in the Vowel chart in Figure 3-1 in section No. 3 (/i/, /y/, /e/, /o/, /a/). Like the consonant IPA symbols, most have very different sounds than when these symbols are used as letters to spell. For example, the IPA symbol /i/ is the "ee" sound of the word "ch*ee*se," and the IPA symbol /e/ is the "ay" sound of the word "b*a*it." Because English spelling doesn't reliably indicate speech sounds, the best way to master the IPA is to

go directly to flash cards and match word sound with IPA symbol. (Refer to the later section "Why the IPA Trumps Spelling" for more information.)

Greek alphabet symbols

The IPA also contains some Greek alphabet symbols. If you're familiar with Greek campus organizations, you may recognize some of them. For instance, consonant symbols include *phi* /ɸ/, *beta* /β/, *theta* /θ/, and gamma /ɣ/. Of these symbols, you find /θ/ in the English words "*th*ing," "au*th*or," and "wor*th*." Among the vowels, you can find *upsilon* /ʊ/ and *epsilon* /ɛ/. You find these sounds in the words "p*u*t" and "b*e*t."

Made-up symbols

The majority of the IPA symbols are made-up characters. They're symbols that have been flipped upside-down or sideways, or they have had hooks or curlicues stuck on their tops, bottoms, or sides. For example, the velar nasal stop consonant, "eng" (IPA character /ŋ/), consists of a long, curled right arm stuck onto a Latin "n." Don't you wish you could have been around when some of these characters were created?

The IPA also has some made-up vowel characters, at least for English speakers. For instance, the IPA mid-front rounded vowel is transcribed /ø/. This is a (lip) rounded version of the vowel /e/, found in Swedish. It sounds like saying the word "bait" while sticking your lips out, causing a lowered sound quality. This symbol resembles an "o" with a line slashed through it.

Another famous made-up vowel is the IPA mid-central vowel, /ə/, *schwa.* This character represents the unstressed sound "uh," as in "th*e*" and "*a*nother."

Tuning In to the IPA

The IPA is broken down into six different parts, which I refer to as charts. Each chart represents different aspects of speech sound classification. Refer to Figure 3-1 to see the different charts. In the following sections, I take a closer look and describe them in greater detail.

Featuring the consonants

The top two charts of the IPA in Figure 3-1 represent the consonants of the world's languages. *Consonants* are sounds made by partially or wholly blocking the oral airway during speech. The large chart (section No. 1) shows 59 different symbols listed in columns by place of articulation and in rows by manner of articulation. Wherever applicable, voiceless and voiced pairs of

sounds (such as /f/ and /v/) are listed side by side, with the voiceless symbol on the left and the voiced symbol on the right.

Because every IPA symbol is uniquely defined by its voicing, place, and manner (see Chapters 2 and 5 for more information), you're now ready to have some fun (and of course impress your friends and family!) by reading off the features for each symbol from the chart. Let me start you off. In the top left box, you can see that /p/ is a voiceless, bilabial plosive. Looking down the next column to the right, you see that /v/ is a voiced, labiodental fricative.

Are you confused and not sure how I came up with these descriptions? Just follow these steps to get them:

1. **Look up to the top of the column to get the consonant's place of articulation.**
2. **Look look to the left side of the row to get the consonant's manner of articulation.**

   If the character is on the left side of the cell, it's voiceless, otherwise it's voiced. If a character is in the middle (by itself), it's voiced.
3. **Put it all together and you have the consonant's voicing, place, and manner of articulation.**

Now it's your turn. Name the voicing, place, and manner of the /h/ symbol in the column at the far right. Yes, /h/ is a voiceless, glottal fricative. Congratulations, you can now cruise to any part of the consonant chart and extract this kind of information. You need this important skill as you work on phonological rules (see Chapters 8 and 9).

Accounting for clicks

The second chart in Figure 3-1 (labeled No. 2) is for sounds produced very differently than in English. When these sounds are produced, air doesn't flow outward from the lungs, as is the case for most language sounds. Instead, air may be briefly moved from the larynx or the mouth. This chart covers the fascinating consonants of Zulu, the sucking-in sounds of Sindhi, and the popping sounds of Quechua, to name a few. Chapter 12 and the multimedia material (located at `www.dummies.com/go/phoneticsfd`) give you some more exposure to these sounds.

Going round the vowel chart

The third chart in Figure 3-1 (labeled No. 3) is called a *vowel quadrilateral,* a physical layout of vowels as produced in the mouth (refer to Figure 3-2 for a better idea what this looks like). In this chart, vowels are represented by how close the tongue is held to the top of the mouth, also known as being *high.* In contrast, the vowel may be produced with an open vocal tract, also known as placing

the tongue *low.* In terms of horizontal direction, the tongue can be described as positioned at the *front, central,* or *back* part of the mouth. Where the symbols are paired, the rightmost symbol is produced with the lips rounded (or protruded). Lip rounding has the effect of giving the vowel a lowered, rather hollow sound.

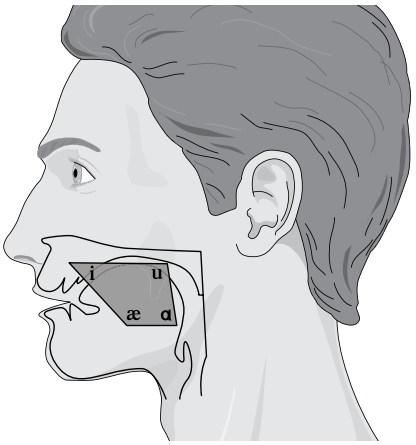


Figure 3-2: Vowel quadrilateral superimposed on a person's vocal tract.

Illustration by Wiley, Composition Services Graphics

Marking details with diacritics

The next chart I focus on in Figure 3-1 addresses the diacritics. (I skip over the chart called "Other symbols," which is a very specialized section.) *Diacritics* (in Chart 4, labeled No. 4) are small helper marks made through or near a phonetic character to critically alter its value. For instance, if you look at the top-left box of this chart, you can see that a small circle, [ₒ], placed under any IPA character, indicates that the sound is produced with a voiceless quality. In other words, if you need to transcribe a normally voiced sound, such as /n/ or /d/ that was produced as voiceless, you can use the diacritic [ₒ].

Stressing and breaking up with suprasegmentals

The fifth chart in Figure 3-1 (labeled No. 5), called *suprasegmentals*, lists the IPA symbols used to describe syllables and words, that is, chunks of speech larger than individual consonants and vowels. This chart includes ways of marking stress, length, intonation, and syllable breaks. For example, the IPA indicates primary stress by placing a small vertical mark in front of the syllable, like this for the word "syllable" /ˈsɪləbəl/. Here, the IPA is different than some books and dictionaries that underline or bold the stressed syllable (like this: **syll**able or *syll*able). I describe this level of phonetics in more detail in Chapter 10.

What are those other symbols?

The IPA section called "other symbols" is designed to cover sounds that don't quite fit in elsewhere. Some of the sounds are produced with two simultaneous constrictions in the vocal tract and thus can't be easily placed in the first section of the IPA. These double-articulations include /w/, /ʍ/, /ɥ/, and /ɧ/. Other sounds have special combinations of manner features that require them being singled out for special designation (/ɕ/, /ʑ/, and /ɭ/). Three sounds in this group (/ʜ/, /ʢ/, and /ʡ/) are produced at the lowest section of the vocal tract, the epiglottis. I provide more information on all of these other sounds in Chapter 16.

Beginning phonetics students are sometimes mortified by the fact that the chart has so many diacritics. You don't have to panic because you only need a small subset to transcribe English. After you begin to see how the diacritic system works, figuring out new characters becomes easier.

Touching on tone languages

The sixth part of Figure 3-1 (labeled No. 6) details special symbols needed for languages known as *tone languages* (such as Vietnamese, Mandarin, Yoruba, or Igbo) in which the *pitch* (high versus low sound) of different syllables and words alter the meaning. This concept may seem odd to monolingual English speakers, because English doesn't have such a system. For example, saying a word in a high squeaky voice versus saying the same word in a much lower voice doesn't change the meaning. However, English-speakers are in the minority, because most of the people of the world speak tone languages. The IPA has a uniform system to mark these tones in terms of their height *level* (from extra low to extra high) and their *contour* (rising, falling, rising-falling, and so forth). Chapter 15 describes tone languages in greater detail.

Sounding Out English in the IPA

The best way to familiarize yourself with the IPA is to practice the different sounds. Practicing can help you understand how these sounds differ and why the IPA chart is organized as it is. Speaking and hearing the sounds can also help you remember them. These sections explain how to make the sounds for the different English IPA sounds.

Cruising the English consonants

Consonants are the first place to start when sounding out the English symbols using the IPA. Figure 3-3 shows the consonants of English.

Manner	Voicing		Place of articulation							
	Voiced (+)	Voiceless (–)	Bilabial	Labio-Dental	Dental	Alveolar	Palato-Alveolar	Palatal	Velar	Glottal
Stop (nasal)	+		m			n			ŋ	
Stop (oral)		–	p			t			k	
Stop (oral)	+		b			d			g	
Fricative		–		f	θ	s	ʃ			h
Fricative	+			v	ð	z	ʒ			
Affricate		–					tʃ			
Affricate	+						dʒ			
Approximant		–	ʍ						ʍ	
Approximant	+		w			ɹ		j	w	
(lateral)	+					l			ɬ	

Figure 3-3: The consonants of English.

To know how to identify one IPA symbol from another, focus on working with a minimal pair. A *minimal pair* is when two words differ by only one meaningful sound. For example, /bæt/ and /bit/ ("bat" and "beet"), or /bæt/ and /bæd/ ("bat" and "bad"). Minimal pairs help people identify *phonemes*, the smallest unit of sound that changes meaning in language. If you become stuck in hearing a particular sound (such as /ŋ/), you may form minimal pair contrasts (such as /sɪn/ and /sɪŋ/ ("sin" and "sing"), to make things clearer.

Here I work through Figure 3-3, column by column. The first column, /m/, /p/, and /b/ are a cinch — they sound like they're spelled in English, as in "*m*at," "*p*at," and "*b*at." All three of these consonants are *stops* (sounds made by blocking air in the oral cavity), the first being nasal, and the last two being oral. Notice at the bottom of the bilabial column you also find symbols /w/ and /ʍ/ — that are also placed in the velar columns. The sounds /w/ and /ʍ/ (voiced and voiceless) are considered *labiovelar*, that is articulations made simultaneously at the labial and velar places of articulation. Such articulations are called *double articulations* and are relatively complex (notice, for example, that young children acquire /w/ sounds relatively late in acquisition).

You make the /w/ sounds with your lips puckered and the tongue held toward the back of your mouth, as in "*w*et" or "*W*illiam." To get a better sense, try to say "wet" without letting your lips go forward — or while holding your tongue tip against your teeth to keep your tongue forward in your mouth. (Doing so is darn near impossible.) Because these double articulated sounds are awkward to fit into the consonant place of articulation chart, they're more typically listed in the Other Sounds section of the IPA. (Refer to Chapter 16 for more information.)

The sound /ʍ/ is like a /w/, but without voicing. Instead of "*w*itch," sounds with /ʍ/ rather sound like "*hw*itch." In fact, at one point the IPA used the symbol "hw" instead of /ʍ/. (I still don't know why they switched!) Some speakers of American English alternate between /w/ and /ʍ/ in expressions such as "*Wh*ich *w*itch is *wh*ich?" (with the middle "witch" being voiced and the others not). If these examples work for you, super! If not, listen to the examples listed in the bonus multimedia material at www.dummies.com/go/phoneticsfd.

Moving to the next column, the labiodentals /f/ and /v/ should also be easy to transcribe. You can find the voiceless consonant /f/ in words such as "*f*ree," "*f*ire," "*ph*one," and "enou*gh*." You can find the voiced labiodentals fricative in "*v*ibe," "ri*v*er," and "Da*v*e."

Students often mix up the dental fricatives /θ/ and /ð/. You can find the voiceless /θ/ in words, such as "*th*igh," "*th*ick," "me*th*od," and "ba*th*." Meanwhile, you can find the voiced fricative /ð/ in words, such as "*th*ose," "*th*is," "la*th*er," "bro*th*er," "la*th*e," and "brea*th*e." You can always sneak your hand up over your larynx (to the Adam's apple), and if you feel a buzz, it's the voiced /ð/.

When you're discovering and mastering new IPA sounds and symbols, I suggest you try them out in all *contexts* (positions in a word) — that is, the beginning, middle, and end. These positions are called word *initial*, *medial*, and *final*. Here are a couple examples:

IPA Symbol	*Initial*	*Medial*	*Final*
/p/	*p*at	a*pp*ea*r*	ri*p*
/f/	*f*in	a*f*raid	shea*f*

Some sounds can't appear in all three positions. For example, the velar nasal consonant /ŋ/ can't begin a word in English. Also, /t/ and /d/ sometimes become a tap in medial position. A *tap* is a very rapid stop sound made by touching one articulator against another, such as the very short "t" sound in "Be*tt*y." Refer to Chapter 9 for more information on these rules.

Acing the alveolar symbols

Many consonant sounds are made at that handy-dandy bump at the roof of your mouth, the *alveolar ridge*. These sounds include /t/, /d/, /n/, /s/, /z/, /ɹ/, and /l/. I describe these sounds in the following list.

- /t/ **and** /d/**:** The case of /t/ and /d/ is interesting. These sounds are pretty straightforward in most positions of American English. Thus, you can find /t/ in "*t*ick," "s*t*eel," and "pi*t*," and you can find /d/ in words, such as "*d*ome," "ca*d*," "*d*rip," and "lov*ed*." However, in *medial* position (the middle of a word), American English has a tendency to change a regular /t/ or /d/ into something called a *tap* or *flap*, which means an articulator rapidly moves against another under the force of the airstream, without enough time to build up any kind of burst, such that it sounds like a fully formed stop consonant. For example, notice that the /t/ in "Be*tt*y" isn't the same /t/ as in "be*t*" — it sounds something like a cross between a /t/ and a /d/ — a short, voiced event. Chapter 9 discusses in great depth the cases when this sound happens.
- /n/**:** Some sounds, such as /n/, are easy for beginning transcribers to work with because their sounds are easy to spot. You find /n/ in the words "*n*ice," "pa*n*," and "ho*n*or."

- **/s/ and /z/:** The fricatives are also relatively straightforward, as in /s/ found in "*s*ail," "ri*c*e," "re*c*eipt," and "fit*s*," and /z/ found in "*z*ipper," "fi*zz*," and "run*s*." But did you notice you can be fooled by spelling, as in "runs" which is spelled with an "s" but actually has a /z/ sound?
- **/ɹ/ and /l/:** These are two additional consonants made at the alveolar place of articulation. *Approximants* are sounds made by bringing the articulators together close enough to shape airflow but not so close that air is stopped or that friction is caused (check out Chapter 6). You can find the consonant /ɹ/ in the words "*r*ice," "ca*r*een," and "c*r*oak." Notice that this IPA symbol is like the letter "r," except turned upside down, because the right-side up IPA symbol, /r/, indicates a trilled (rolled) "r", as in the Spanish word "burro." Some phonetics textbooks incorrectly let you get away with transcribing English using /r/ instead of /ɹ/, but I recommend forming good habits and using /ɹ/ whenever possible!

Saying, "I'm chilling with phonetics" isn't completely inaccurate, because sucking in cool air while holding the mouth position for any given consonant is an effective way to feel where your articulators are. Try it with the *lateral* alveolar consonant, /l/. Make the /l/ of the syllable "la," and hold the /l/ while sucking in air through your mouth. You should feel cool air around the sides of your tongue, showing that this is a *lateral* (made with the sides) sound. You may also notice a kind of Daffy Duck-like slurpy sound quality when you attempt it.

In the same column, under /ɹ/ you can see the symbol /l/. You can make a lateral sound by passing air around the sides of the tongue, which is different than most sounds, which are *central,* with airflow passing through the middle of the vocal tract. The consonant /l/ is another interesting case that occupies two columns in the consonant chart for English — you can also find it in the velar column.

There are actually two slightly different flavors of /l/:

- **Light /l/:** This one is produced at the alveolar ridge. You can always find the light l at the beginning of a syllable. It has a higher sounding pitch. Some examples include "*l*ight," "*l*eaf," and "*l*oad."
- **Dark /ɬ/:** This one is produced at the back of the tongue. The dark l, also called *velarized,* is marked with a tilde diacritic /~/ through its middle. The dark l occurs at the end of a syllable and sounds lower in pitch. Some examples include "waff*le*," "fu*ll*," and "ca*ll*."

Pulling back to the palate: Alveolars and palatals

The English *palato-alveolar* (or *post-alveolar*) consonants consist of two manners of articulation:

- **Fricatives:** The *fricatives* are represented by the voiceless character "esh" or "long s." Words with this sound include "*sh*eep," "na*ti*on," "mi*ss*ion," "wa*sh*," and "*s*ure." The voiced counterpart, "ezh" or "long z," /ʒ/ is rarer in English, including words, such as "mea*s*ure," "lei*s*ure," "rou*g*e," and "deri*si*on." There are almost no cases of word-initial /ʒ/ sounds (except Zsa Zsa Gabor).
- **Affricates:** The *affricates* /ʧ/ and /ʤ/ are sounds that begin abruptly and then continue on a bit in hissy frication. Some examples of the voiceless /ʧ/ include "*ch*ip," "*ch*ocolate," "fea*tu*re," and "wa*tch*." When a person voices this sound, it's /ʤ/, as in "*Ge*orge," "re*gi*on," "mi*dg*et," and "ju*dg*e." Again, if you have any problems knowing which is voiced and which is voiceless, reach up and feel your Adam's apple to see whether you're buzzing or not.

The palatal consonant /j/ is interesting. You can find this sound in words, such as "*y*es," "*y*outh," and "*y*ellow." However, it also occurs in the words "*fe*w," "*cu*te," and "*mu*te." To see why, here's a minimal pair: /mut/ versus /mjut/, "moot" versus "mute." You can see that "mute" begins with a palatalized /m/, having a palatal glide /j/ right after it. Slavic languages (like Russian and Polish) use palatalized consonants much more than English; in fact, when teaching English as a second language (ESL) to these speakers, breaking them of this habit can be quite a challenge.

Reaching way back to the velars and the glottis

Three additional stop consonants are in the velar column, the oral stops /k/ and /g/, and the nasal stop /ŋ/. Examples of /k/ include "*C*arl," "s*k*in," "e*xc*ess," and "ra*ck*." Examples of /g/ include "*g*irl," "ag*g*ravate," and "fo*g*." Notice that /g/ corresponds with what some call "hard g," not a "soft g."

The last sound in the chart is what one might call "way down there." That is, the glottal fricative, /h/. Your *glottis* is simply a hole or space between your vocal folds in your throat. When you cause air to hiss there, you get an "h" sound, as in "*h*ello," "*h*ot," "*wh*o," and "a*h*a!" In Chapter 2, I discuss making a stop with your glottis (a *glottal stop*, /ʔ/) — however, you don't freely use this sound to make words in English; instead, it alternates and only appears under certain conditions. As such, glottal stop and flap are special sounds (called *allophones*) that aren't included in the main chart.

Visualizing the GAE vowels

English vowels are more difficult to describe than English consonants because they're produced with less precision of tongue positioning. Vowels differ systematically across major forms of English (such as American and British).

Between these two major dialects, one major difference is the presence or absence of *rhotacized* (r-colored) vowels. Whereas most GAE speakers would pronounce "brother" as /ˈbɹʌðɚ/, most British speakers pronounce it as /ˈbɹʌðə/. The difference is whether the final vowel has an r-like quality (such as /ɚ/) or not (/ə/). Refer to Chapters 7 and 18 for more information about American and British vowel differences. Vowels typically differ across the dialects within any given type of English. For example, within American English think of the difference between a talker from New York City and one from Atlanta, Georgia. In British English, one would expect differences between speakers from London (in the south) and Liverpool (in the north).

Figure 3-4 is a chart of the vowels most commonly found in General American English (GAE).

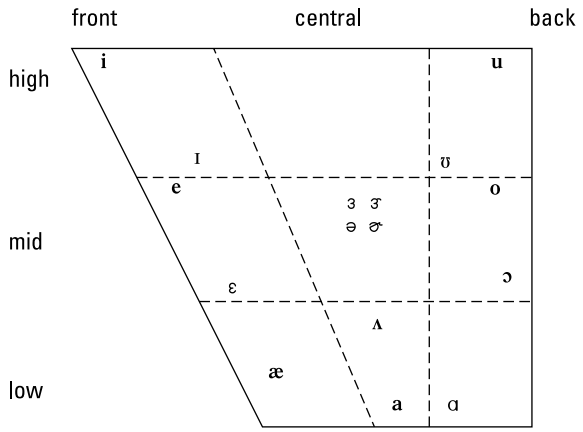


Figure 3-4: Vowels of General American English.

In Figure 3-4, I use the terms *high* and *low* in place of IPA *close* and *open.* To keep things simple, I also use "h_d" words, as examples to capture the typical vowels produced by speakers of General American English.

Starting with the front vowels, say "heed," "hid," "hayed," "head," and "had." These five words include examples of the front vowel series, from high to low. You can find the symbol /i/, lower case "i," in the words "fl*ee*ce," "p*ea*," and "k*ey*." A vowel slightly lower and more central is /ɪ/, "small capital I", as in the words "th*i*ck," "t*i*p," "*i*llustrate," and "r*i*d."

Say that you're a speaker of English as a second language (ESL) and come from a language like Spanish that has /a/, /i/, /u/, /e/, and /o/ vowels (but not /æ/, /ɪ/, /ʊ/, /ɛ/, and /ɔ/ vowels). I discuss more about these vowel differences in Chapter 7. For now, you may need to work a bit extra to be able to identify these English sounds. Using minimal pairs is a good way to sharpen up your ears!

My Fair Lady: Famous phonetics story with an important message

A world-famous story dealing with phonetics is the musical *My Fair Lady,* based on the play, *Pygmalion*, by George Bernard Shaw. In this drama, a British phonetician, Henry Higgins, teaches a lower-class flower girl, Eliza Doolittle, to switch from her Cockney accent to proper English. This saga is a satire of the British class system, a love story, and a taste of phonetics, all in one.

The goal of changing someone's accent is called *prescription* (judging what is correct). Today, linguists and phoneticians spend much less time *prescribing* how people should sound and more time *describing* how different languages and dialects do sound. There is still a market for foreign accent reduction, although a bit different than at the time of Eliza Doolittle. Also, as wonderful as the song is, clients aren't necessarily required to sing "The rain in Spain stays mainly in the plain." Refer to Chapter 1 for more information on prescribing and describing.

The symbol /e/ is a mid-front vowel, as in "s*ai*l," "*a*pe," and "l*a*zy. "You can find the symbol /ɛ/, *epsilon*, in the words "l*e*t," "sw*ea*ter," "t*e*ll," and "t*e*n." The low-front vowel, /æ/ is called *ash.* Phoneticians introduced this Old English Latin character into the IPA. To write an ash, follow the instructions in Figure 3-5.

ɪ			
ɛ			
æ			
ə		ŋ	
ɚ		ʃ	
ʌ		ʒ	
ɝ		θ	
ʊ		ð	
ɔ		ɾ	
ɑ		ʔ	

Figure 3-5: How to draw some of the common made-up IPA symbols.

To master the symbols for the GAE back vowels, say "who'd," "hood," "hoed," "hawed" (as in "hemmed and hawed"), and "hospital." (You can also say "hod," but few people know what a hod [coal scuttle] is anymore.) These words represent the back vowel series /u/, /ʊ/, /o/, /ɔ/, and /ɑ/, which I discuss here with some examples:

- **/u/:** You can find this high back vowel in the words "bl*ue*," "c*oo*l," and "ref*u*sal."
- **/ʊ/:** This symbol has a Greek name, *upsilon,* and you form it by taking a lower case u and placing small handles on it. You can find this sound in "p*u*ll," "b*oo*k," and "w*ou*ld."
- **/o/:** The mid-back vowel can sometimes sound pretty much like it's spelled. You can find it in words, such as "t*oe*," "g*o*," "*ow*n," and "mel*o*dious."
- **/ɔ/:** This mid-low vowel is called *open-o* and is written like drawing a "c" backwards. You can find this vowel in the words "s*aw*," "b*a*ll," "*awe*," and "l*aw*," like most Americans pronounce.
- **/ɑ/:** You can find this low-back vowel, referred to as *script a,* in the words "f*a*ther," "psych*o*logy," and "h*o*nor."

You may have noticed a different flavor of the vowel "a," in Figure 3-4, found slightly fronted to script a. This IPA /a/, "lower case a," is used to indicate the beginning of the English diphthongs /aɪ/ and /aʊ/, as in "m*i*le" and "l*ou*d."

Why the IPA Trumps Spelling

When it comes to explaining language sounds, English spelling doesn't have the power or the precision to deal with the challenge because there is a loose relationship between English letters and language sounds. Therefore, a given sound can be spelled many different ways. Here are some famous examples:

- The word "ghoti" could logically be pronounced like "fish." That would be the "gh" of "enou*gh*," the "o" of "w*o*men," and the "ti" of "na*ti*on." Playwright and phonetician George Bernard Shaw pointed out this example.
- The vowel sound in the word "*ei*ght" (transcribed with the symbol /e/ in IPA) can be spelled "ay," "ea," "au," "ai," "ey," and "a (consonant) e" in English. If you don't believe this, say the words "day," "break," "gauge," "jail," "they," and "date."
- Many languages have sounds that can't be easily spelled. For instance, Zulu and Xhosa have a consonant that sounds like the clicking noise you make when encouraging a horse ("tsk-tsk") and another consonant that sounds like a quick kiss.
- Most world languages convey meaning by having some syllables sound higher in pitch than other syllables.

Chapter 4

Producing Speech: The How-To

In This Chapter

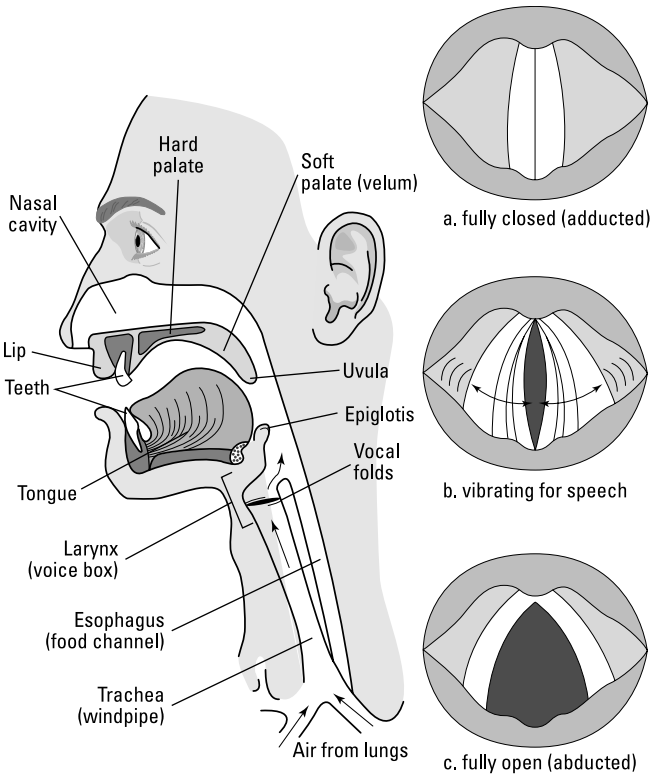
- Knowing how your body shapes sounds
- Getting a grounding in speech physiology
- Looking closer at speech production problems
- Seeing how scientists solve speech challenges

Understanding not only what parts of your body are involved in making speech is important, but also which mechanical and physiological processes are involved. That is, how do you produce speech? This chapter gives more information about the source of speech, addressing how high and low voices are produced, and how people shout, sing, and whisper. I provide many more details about how sounds are shaped, so that you can better understand the acoustics of speech (which I discuss more in Chapter 12). At the end of the chapter, you can compare your own experience of producing speech with current models of speech production, including those based on speech gestures and neural simulations.

Focusing on the Source: The Vocal Folds

To have a better understanding of the source of the buzz for voiced sounds, you need to take a closer look at the vocal folds and the larynx. The *vocal folds* (also known as *vocal cords*) are small, muscular flaps located in your throat that allow you to speak, while the *larynx* (also known as the *voice box*) is the structure that houses the vocal folds. Refer to Chapter 2 for more background about general speech anatomy. For this discussion, Figure 4-1 gives you some details about the vocal folds and larynx.

The following sections explain some characteristics of vocal folds and how they work, including what they do during regular speech, whispering, loud speech, and singing.



Wiley, Composition Services Graphics

Figure 4-1: A diagram of the vocal folds in the larynx: fully closed (adducted) (a), vibrating for speech (b), and fully opened (abducted) (c).

Examining the Bernoulli principle

Daniel Bernoulli (1700–1782) was a prolific scientist and member of the famous Bernoulli family (originally of Antwerp, then in the Spanish Netherlands), many of whom were scientists, mathematicians, and artists. Bernoulli is perhaps most known for his principle, which states that fluids in an area moving faster than the surrounding area possess less pressure. That is, the faster moving the fluid is, the lower the pressure. *Fluid* generally includes gases, such as air.

The Bernoulli principle can explain why if you're walking along the side of a road and a giant truck goes roaring past, you may feel sucked into the middle of the road in its wake. The truck creates a high-speed blast of fluid (air) pressure next to you, lowering the pressure. You're then pulled into that low-pressure minimum.

You can test this principle by taking two very light, aluminum cola cans and placing them about one to two inches apart on a bed of soda straws. If you then blow sharply between the cans using a straw (imitating the force of the giant truck), you can watch the cans be sucked inward into the low-pressure gradient.

The Bernoulli principle regulates vocal fold adduction by creating a low-pressure zone that draws in the vocal folds. The forces draw together the vocal folds and the tracheal pressure pushes them apart. In this manner, the pulse chain of vocalization is sustained.

Identifying the attributes of folds

The vocal folds are an important part of your body that can't be seen without a special instrument. Located deep in your throat, these small muscular flaps provide the buzzing source needed for voiced speech. Check out these important characteristics about vocal folds:

- The male vocal folds are between 17 and 25 millimeters long.
- The female vocal folds are between 12.5 and 17.5 millimeters long.
- The vocal folds are pearly white (because of scant blood circulation).
- The vocal folds are muscle (called the *thyroartyenoid* or *vocalis*), surrounded by a protective layer of mucous membrane.
- When the vocal fold muscles tighten, their vibratory properties change, raising the pitch.
- A person can possibly speak with just one vocal fold; however, people sound different than before. For example, Jack Klugman (who played Oscar in *The Odd Couple*) had his right vocal fold surgically removed due to laryngeal cancer. To hear samples of his speech before and after, go to: `minnesota.publicradio.org/display/web/2005/10/07_klugman/` and `www.npr.org/templates/story/story.php?storyId=5226119`.

Pulsating: Vocal folds at work

In order for the vocal folds to create speech, several steps must take place in the right order. Follow along with these steps and refer to Figure 4-2:

1. **The vocal folds *adduct* (come together) enough that air pressure builds up beneath the larynx, creating tracheal pressure.**
2. **The force of the ongoing airstream *abducts* (brings away from each other) the vocal folds.**

   To keep straight the directions of abducting and adducting, remember that the glottis is basically a hole (or an absence). Thus, abducting the glottis creates a space, where as adducting means bringing the vocal folds together.
3. **The ongoing airstream also keeps the vocal folds partially adducted (closed) because of the Bernoulli principle.**

   The Bernoulli principle states that fast moving fluids (gases) create a sort of vacuum that may draw objects into its wake. Refer to the nearby sidebar for more information about this property.

4. **The vocal folds flutter, with the bottom part of each fold leading the top part.**
5. **Under the right conditions, this rhythmic pattern continues, creating glottal *pulses* of air, a series of steady puffs of sound waves.**

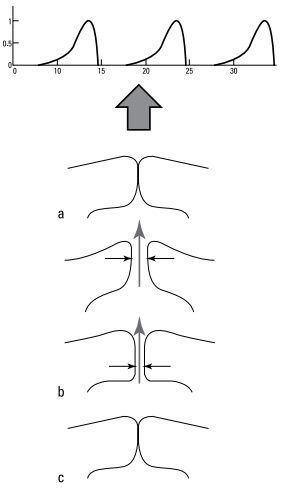


Figure 4-2: How the vocal folds produce voicing for speech.

Wiley, Composition Services Graphics

The faster the pulses come means the higher the *fundamental frequency* (rate of pulses per second). Fundamental frequency is heard as *pitch* (how high or low a sound appears to be). The way your body regulates fundamental frequency is to adjust the length and tension of the vocal folds. A muscle called the *cricothyroid* raises pitch by rocking the thyroid cartilage against the *cricoid* cartilage (which is ringlike), elongating the vocal folds. When the vocal folds are stretched thin, they vibrate more rapidly. For instance, strong contraction of the cricothyroid muscle gives voice a falsetto register (like the singer Tiny Tim).

If you wish to try your own cricothyroid rocking experiment, put your thumb and forefinger over your cricothyroid region (see Figure 4-3) and sustain the vowel /i/. If you jiggle your fingers in and out (not too hard), you can cause rocking on the cricothyroid joint and create a slight pitch flutter.

The vibrating vocal folds are commonly viewed using an instrument called an *endoscope*, a device that uses fiber optics to take video images during speech and breathing. Endoscopy images can either be taken using a rigid wand placed through the mouth at the back of the throat (*rigid endoscopy*) or via a thin, flexible light-pipe fed through the nostril down just over the larynx (*flexible endoscopy*). Strobe light can be pulsed at different speeds to freeze-frame the beating vocal folds, resulting in stunning images. To see videos of the vocal folds during speech taken at different fundamental frequencies of phonation, see `http://voicedoctor.net/videos/stroboscopy-rigid-normal-female-vocal-cords-glide` and `www.youtube.com/watch?v=M9FEVUa5YXI`.

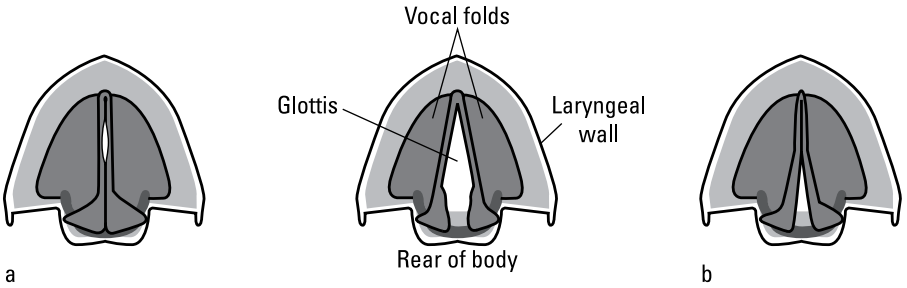
Wiley, Composition Services Graphics

Figure 4-3: Two fingers placed over cricothyroid region for rocking experiment.

Here are some important facts about the buzzing you do for speech:

- During speech, roughly half of the consonants you produce and all of the vowels are voiced.
- The vocal folds are drawn tight.
- There is more of an opening at the posterior portion than the front.
- Men's vocal folds vibrate on average 120 times per second.
- Women and children's vocal folds vibrate at a higher frequency than those of men (due to smaller size). On average, women's vocal folds beat 220 times per second, while children's beat around 270 cycles per second.

Figure 4-4 shows the vocal folds during voiced speech (Figure 4-4a) and whispered speech (Figure 4-4b). These sections also examine what your vocal folds specifically do when you yell and sing.



Wiley, Composition Services Graphics

Figure 4-4: What a glottis does during voiced speech and whispering.

Whispering

Opening the glottis somewhat, which allows air to flow out while creating friction, creates whispered speech (refer to Figure 4-4b). This process is similar to what creates the voiceless fricative consonant "h" as in "*h*ello" (/h/ in IPA).

There is no language where people whisper instead of talking because whispered speech isn't as understandable as spoken language; it's simply not as loud or clear. However some languages mix whispering with regular voiced speech in a special way to produce a distinctive feature called *breathiness* that can change meaning. For instance, if you're visiting Gujurat, India, and wish to visit a "palace" (a word pronounced in Gujurati with breathy voice), you don't want to use the word for "dirt," which is the same word pronounced without breathiness. Refer to Chapter 15 for more information.

Talking loudly

Your breathing system (including your lungs and trachea), your larynx, and the neck, nose, and throat regulate speech volume. The more air is passed through the glottis (for instance, at higher tracheal pressures), the higher the air pressure of the voice. Raising the resistance of the upper airway, by reducing the size of the glottis and not letting air escape needlessly, can also increase the pressure. In addition, opening the pharynx and oral cavity to greater air volumes increases resonance and allows sound to flow less impeded. This opening of the pharynx and oral cavity can include elevating the velum, lowering the jaw/tongue, and opening the mouth.

Being loud and proud

Who is the loudest person in the world? When scientists chart how loud someone is, they usually measure the lowest and highest sound pressure that can be registered at a certain fundamental frequency. Researchers have converted this sound to a likely heard (perceptible) value with a formula. They chart these sounds with sound pressure levels in decibels (dB) on the vertical axis and fundamental frequency on the horizontal axis. In general, people tend to produce lower sound pressure levels with lower frequency sounds and higher pressures with higher frequency sounds. The cutoff for the highest sounds appears to be around 109 dB.

However, some reports of various contests worldwide claim individuals have topped the norm. For years, the record belonged to Simon Robinson, who reached an epic 128 dBs at a distance of 8 feet, 2 inches during a competition in Adelaide, Australia. More recently, Jill Drake, a 52-year-old teaching assistant in Kent, England, broke the record with a 129 dB shout, approximately the same level as a jackhammer.

Children may take time to develop the *sensorimotor* (body-sensing) systems necessary to regulate voice volume during speech. For instance, child language researchers report (anecdotally) that young children can have difficulty in adjusting their volume in speaking tasks; they tend to be quiet or loud without gradations in between, which may also explain why children have trouble speaking with their "inside" voice.

Too much loud speech can damage the vocal folds; voice clinicians work on a daily basis by assigning warm-up exercises, periods of rest, hydration, and other relaxation tips to help reduce stress and strain on the professional voice.

Singing

Singing is a part of musical traditions throughout the world. When you listen to other languages, they can sometimes sound melodic or a bit like singing. However, in other ways the sounds of a foreign language are clearly different from the sounds of someone singing. Although speech and singing research show the two are closely linked, they do have interesting differences in terms of vocal production.

English speakers make more voiced sounds during singing (around 90 percent) than during speech (around 40 percent). People usually sing from a pre-defined score or memorized body of material, with the goal of more than just the communication of words but also to convey emotion, intent, and a certain sound quality. As such, sung *articulatory gestures* (lip, jaw, and tongue movements) are generally exaggerated, compared to everyday speech.

An interesting clue about the kind of information people can include in the sung voice comes from studying the voices of opera singers. Johan Sundberg, a professor at the University of London, has conducted extensive research into the acoustics of singing. In a number of famous studies, he developed the idea of the *singing formant,* an additional resonant peak (at around 4 to 5 kHz), which results from lowering the larynx. This peak has the effect of making the sung voice stand out from a background of orchestral music. See Chapter 12 for more information on formants and resonant peaks.

Other kinds of sung voice exist besides opera, including gravelly or rough voices, used in genres such as folk, blues, and rock. In ongoing studies, researchers are investigating what is at the core of these types of sung voices, even going so far as to study ugly voice (that may make bad singers *not* sound good).

The quest to replace vocal folds

The precious half-inch to three-quarters inch strips of muscular tissue in your throat that allow you to make voiced sounds usually last a lifetime. But what if something goes wrong? Cancers, infections, and surgical complications, as well as stomach acids, reflux, and general wear and tear (especially in professional singers) can cause these tissues to malfunction. Doctors usually noninvasively treat minor problems, whereas in cases of paralysis from surgical complications or for individuals with *laryngectomies* (whole or partial vocal fold removal) scientists are looking at the following ways of replacement and repair.

- **Vibrating gels:** One exciting avenue focuses on developing gels that vibrate with approximately the same characteristics as the vocal folds. For instance, Dr. Robert Langer and colleagues at Harvard University and MIT are working on gel-like materials that vibrate at around 200 Hz (similar to a female voice) when stimulated with the same amount of air pressure that would normally be exerted at a human glottis. Individuals missing vocal fold function would receive gel injections to boost vibrations.
- **Vocal fold augmentations:** In cases of individuals with unilateral paralysis or vocal fold atrophy, surgeons are perfecting vocal fold *augmentation* (increasing, enhancing, or enlarging) techniques. For many years, doctors have preferred injecting Teflon into shrunken or missing parts of vocal folds. However, more recently surgeons are using living tissue because it leads to regeneration of vocal fold tissue. In this procedure, surgeons harvest a small amount of fat cells and connective tissue from the patient's own thigh and inject them into the affected vocal fold.
- **Human larynx transplants:** A third exciting avenue is to transplant an entire human larynx for individuals with total laryngectomies. A first transplant was attempted in 1998 at the Cleveland Clinic, restoring the voice of Timothy Hediler after a motorcycle accident. He spoke normally for the first eight years after the transplant, but later he experienced some swelling in his vocal cords that made his voice sound a bit breathy and froggy. Despite that, doctors said his quality of life improved.

  In 2011, surgeons at UC Davis, headed by Dr. Peter Belafsky, transplanted a larynx into 52-year-old Brenda Jensen, who had damaged her vocal cords after repeatedly pulling out a breathing tube while under sedation in the hospital. The operation lasted 18 hours over two days, performed by doctors who had trained two years for the procedure. Surgeons replaced her larynx, windpipe, and thyroid gland with that of a donor who died in an accident. In numerous appearances after the operation, she reported being able to speak with "her own voice."

Recognizing the Fixed Articulators

The bedrock of your speech anatomy is your skull. This includes your teeth, the bony (alveolar) ridge that contains the teeth, and the hard palate, just

behind the teeth. Before examining the moving organs that shape speech (most notably, the tongue), I focus on the key regions where speech sounds are made. This section gives special attention to *compensatory* (or counterbalancing) effects that these fixed structures may have on other parts of your speaking anatomy.

Chomping at the bit: The teeth

You're born with no visible teeth, just tiny indentations. You grow 20 baby teeth by about age 2½ and then shed them and grow a set of about 32 permanent teeth by about 14 to 18 years of age. Besides providing employment for the Tooth Fairy (and your dentist), research shows that your teeth (officially known as *dentition*) may have mixed effects on speech.

A couple ways that teeth can affect speech include the following:

- **Compensatory articulation:** People show compensatory articulation when they speak. *Compensatory articulation* means that a talker can produce a sound in more than one way. If one way of producing a sound isn't possible, another way can be used. Shedding *deciduous teeth* (also referred to as *baby teeth* or *milk teeth*) can cause speech errors, particularly with front vowels and fricatives. However, such complications are usually temporary and people normally overcome them.

  For instance, you ordinarily produce the fricative /s/ by creating a hissing against the alveolar ridge and having the escaping air shaped by your front teeth. However, if you shed your front teeth at age 8, you may hiss with air compressed slightly behind the alveolar ridge, while using a somewhat more lateral escape. This "s" may sound rather funny, but most listeners would get the general idea of what you're saying. Chapter 14 provides more information on compensatory articulation.

- **Jaw position:** A more serious type of effect that the teeth may have on speech is through their indirect effect on jaw position. The teeth and jaw form a relationship called *occlusion*, more commonly known as the *bite type.* In other words, occlusion is the relation between your upper jaw (the *maxilla*) and your lower jaw (the *mandible*). See the section "Clenching and releasing: The jaw" later in this chapter.

Sounds made at the teeth in English include the interdental fricative consonants (voiceless /θ/ and voiced /ð/), as well as the labiodental fricative consonants (voiceless /f/ and voiced /v/). British, South African, Australian, and other varieties of English produce many dental "t" and "d" sounds (see Chapter 18), whereas General American and Canadian English accents use glottal stop /ʔ/, alveolar flap /ɾ/, and alveolar /t/ or /d/.

Imaging the palate: Then and now

Phoneticians have long used the hard palate as a region to take interesting images of speech articulation. Daniel Jones (a key British phonetician who inspired *Pygmalion* and *My Fair Lady*) observed the place of articulation for lingual consonants. First, he had the patient open his mouth. Next, he would blow out a (short) candle and quickly place this in the patient's mouth, blackening the palate with candle soot. He then had the patient articulate, for instance with /ɑsɑ/. Finally, he inserted a mirror and observed. He identified a mark where the patient's wet tongue had touched the candle soot, revealing an image of the place of articulation on the palatal surface.

The technologies of electropalatography (EPG) and electro-optical palatography use a similar approach today. Such patients are taught to read a visual display and then to emulate contact patterns produced by a therapist. In EPG, the speaker wears a custom-made artificial palate (like a dentist's retainer). The artificial palate contains numerous contact-sensitive electrodes that are activated when touched by the tongue. In this manner, phoneticians can use EPG to record patterns of tongue contact during speech, which can be useful for recording consonant production. Scientists have used this technology to learn about speech motor planning and control. For instance, coarticulation of the tongue in different vowel environments has been described with the help of EPG. Clinicians have used EPG to provide real-time speech feedback to a variety of populations, including children with cleft palate, children with Down syndrome, children who are deaf, children with cochlear implants, children with cerebral palsy, adults with Parkinson's disease, and adults with speech apraxia.

Electro-optical palatography is a less common technique where a patient wears an artificial palate that contains optical reflective sensors. These sensors act like tiny video cameras that track the tongue, not by sensing, but by imaging. Electro-optical palatography systems can track not only consonants, but also vowels. Although these systems are still in early stages of development, one day a speech scientist or speech language pathologist may simply ask the patient to pop in a retainer and an image of the tongue, shot from the roof of the patient's mouth, will appear on screen.

Making consonants: The alveolar ridge

Phoneticians are concerned with the upper *alveolar ridge,* the bump on the roof of your mouth between the upper teeth and the hard palate, because it's where many consonants are made. Examples of alveolar consonants in English are, for instance, /t/, /d/, /s/, /z/, /n/, /ɹ/, and /l/ like in the words "*t*oday," "*d*ime," "*s*oap," "*z*oo," "*n*ice," "*r*ose," and "*l*augh." Refer to Chapter 6 for more details.

Aiding eating and talking: The hard palate

The *hard palate* is the front part of the top of your mouth, covering the region in between the arch formed by the upper teeth. It's referred to as hard because of its underlying bones, the skull's *palatine bones*. Take a moment to feel your hard palate — run your tongue along it. It should feel, well, hard. You should also feel ridges on it, called *rugae*. These ridges help move food backwards toward the throat.

The hard palate is an essential part of your body for eating and talking (although not at the same time). English sounds made at the hard palate include /j/, /ʃ/, and /ʒ/, as in "*y*ou," "*sh*ale," and "mea*s*ure."

Palate shape can have an effect on speech. Recent work by Professor Yana Yunusova and colleagues at the University of Toronto have shown that individuals with very high (domed) palates produce very different articulatory patterns for vowels and consonants than individuals with flat and wide palate shapes. Nevertheless, both sets of talkers can produce understandable vowels.

Some individuals have birth defects called *cleft palate.* These disorders result in extreme changes in hard (and soft) palate shape caused by an opening between the mouth and nasal passage. The effect on speech is called *velopharyngeal-nasal dysfunction*, a problem between making oral and nasal closures for speech (refer to the later section, "Eyeing the soft palate and uvula: Velum" for more information).

Locating the hidden hyoid

The *hyoid,* named after the letter *upsilon* because it is u-shaped, is the only bone in the body not connected directly to other bones. It sits right below your tongue and jaw, above the thyroid cartilage and the larynx. Your hyoid isn't really a speech articulator; instead it's an important point of attachment (an anchoring) for the speech muscles of the tongue, larynx, and pharynx to hold onto.

On a somewhat grisly note, the hyoid bone is a telltale sign of strangling used in forensics. When a person is strangled, the hyoid becomes highly compressed and changes shape. This distortion indicates strangling.

Until recently, no ancient hyoid bones were found of human ancestors or related homonids. However, in 1989 archeologists found a Neanderthal specimen in a cave in Kebara, Israel, that had a remarkably modern-looking hyoid. This discovery led some archeologists to conclude that Neanderthal was capable of modern speech and language because this modern hyoid suggested a descended larynx, while other scientists countered that hyoid shape doesn't determine larynx position.

Eyeing the Movable Articulators

A great deal of speech lies in the movement of your articulators. For this reason, I like to refer to the "dance of speech." Speech movements are quick, precise, and fluid — like a good dancer. To speak, you need a plan, but you can't follow it too tightly; instead, the movements are flowing, overlapped, and coordinated. Everything comes together by sticking to a rhythm. These sections focus on those parts of the body that accomplish this amazing speech dance.

Wagging: The tongue

The tongue is the primary moving articulator. In fact, it's quite active in a wide range of activities. The tongue can stick out, pull in, move to the sides and middle, curl, point, lick, flick up and down, bulge, groove, flatten, and do many other things. You use it for eating, drinking, tasting, cleaning the teeth, speaking, and singing (and even kissing).

It's a large mass or muscle tissue; the average length of the human tongue from the *oropharynx* (top part of the throat) to the tip is 10 centimeters (4 inches). The average weight of the adult male tongue is 70 grams, whereas a female's is 60 grams.

The size of a newborn's tongue pretty much fills the oral cavity, with the tongue descending into the pharyngeal cavity with maturation. The tongue develops, along with the rest of the vocal tract, through childhood and reaches its adult size at around age 16.

Although the tongue may look like it's moving really fast, typical speech movements actually aren't as fast as, say a human running. They're on the order of centimeters per second, or around a mile per hour. However, it's the astounding coordination of these tongue movements as sound segments are planned and blended that is hard to fathom.

Researchers continue to study how such movements are planned and produced. Direct study of tongue movement in a number of languages has suggested that much of the variance in tongue shapes falls into two main categories:

- **Front raising:** The tongue moves along a high-front to a low-back axis.
- **Back raising:** The tongue bunches along a high-back to low-front axis.

However, this basic explanation doesn't fit all sounds in all contexts, and researchers are continuing to search for better models to describe the complexity of tongue movement during speech.

Many people make the mistake of underestimating the tongue's size and shape, based on observing their own tongue in a mirror. Doing so is a mistake because the image of one's own tongue only shows the tip (or *apex*) and blade, just a small part of the entire tongue itself. In fact, most of the tongue is humped, which you can't see in a mirror. The tongue, except for a thin covering, is almost entirely muscle. Figure 4-5 shows its structure.

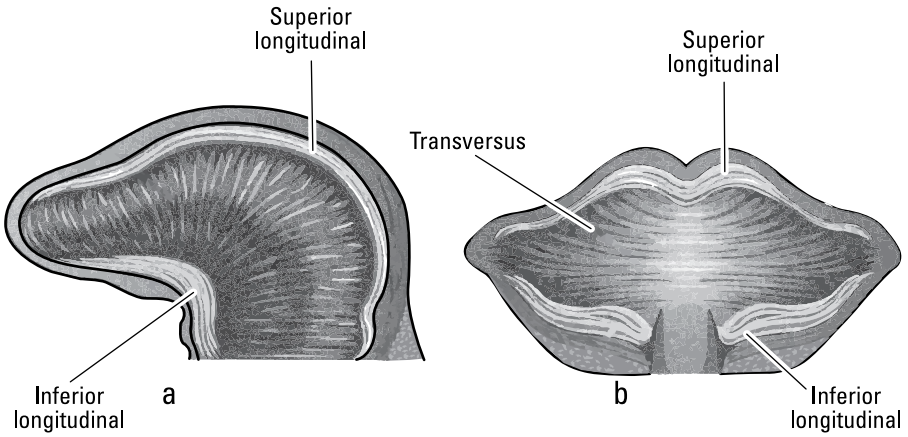


Figure 4-5: A tongue's intrinsic muscles from a side view (a) and front (b) view.

Wiley, Composition Services Graphics

Figure 4-5 shows that the tongue consists of four muscles, called *intrinsic* muscles (inside muscles) that run in different directions. These four muscles are the *superior longitudinal, inferior longitudinal, verticalis,* and *transversus.* When these muscles contract in different combinations, the tongue is capable of numerous shapes.

Extrinsic muscles, which are outside muscles, connect the tongue to other parts of the body. These muscles (refer to Figure 4-6) position the tongue. The extrinsic muscles are the *genioglossus, hyoglossus, styloglossus,* and *palatoglossus*. The names of these muscles can help you understand their functions. For instance, the hyoglossus (which literally means "hyoid to tongue") when contracted pulls the tongue down toward the hyoid bone in the neck, lowering and backing the tongue body.

Your tongue is the one part of your body most like an elephant because the tongue is a *muscular hydrostat*, like an elephant's trunk. A hydrostat is a muscular structure (without bones) that is incompressible and can be used for various purposes. When the tongue extends, it gets skinnier. When it withdraws, it gets fatter. Think elephant trunk, snake tongue, or squid tentacles.

By the way, creating a tongue from scratch isn't easy. To see some of the latest attempts in silicon modeling of the tongue conducted by researchers in Japan, refer to the bonus online Part of Tens chapter at `www.dummies.com/extras/phonetics`.

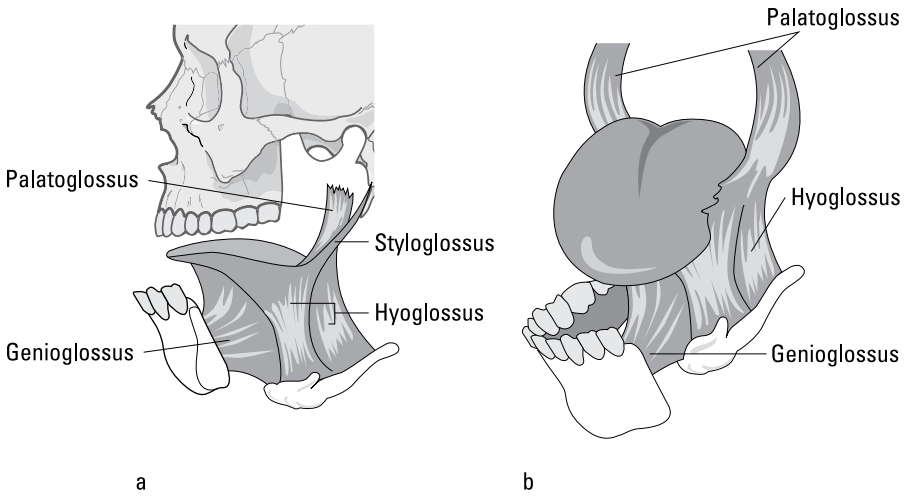


Figure 4-6: The tongue's extrinsic muscles from a side view (a) and an oblique view (b).

Wiley, Composition Services Graphics

More than just for licking: The lips

The lips comprise the *orbicularis oris* muscle, a complex of muscles that originate on the surface of the jaws and insert into the margin of the lip membrane and chin muscles. The lips act to narrow the mouth opening, purse the opening, and pucker the edges. This muscle is also responsible for closing the mouth. The lips act like a sphincter but the lips comprise four different muscle groups, therefore the lips aren't a true sphincter muscle.

In English, the lips are an important place of articulation for the bilabial stop consonants /p/, /b/, and /m/, for the labiodental fricatives /f/ and /v/, and for the labiovelar approximants /w/ and /ʍ/.

Your lips are important in contributing to the characteristics of many vowels in English, for instance — /u/, /ʊ/, /o/, and /ɔ/. When you form these vowels, lip rounding serves as a *descriptive*, but not a *distinctive* feature. That is, when your lips form the features of these four vowels in English, this lip rounding doesn't distinguish these four from any other set that doesn't have lip rounding. Another descriptive feature example is the English vowel /i/, made with lips spread. Acoustically, spread lips have the effect of acting like a horn on the end of a brass instrument, brightening up the sound. In this case, lip spread, not lip rounding, is a descriptive feature for /i/.

In languages with phonemic lip rounding, the planning processes for lip protrusion are generally more extensive and precise than those in English (check out Chapter 6 for more information).

Clenching and releasing: The jaw

The jaw, also known as the mandible, is a part of your body that seems to drive scientists crazy. It is distinct shape-wise from the rest of your body both in terms of its proportions and specific anatomical features.

The jaw keeps its shape as it grows with the body throughout maturation. In fact, it's the only bone in the body to do so. The jaw is a moving articulator that is involved in speech, primarily as a platform for the tongue. Recent studies have also suggested that people can voluntarily control jaw stiffness, which can be useful when producing fine-tuned sounds, such as fricatives, where the tongue must be precisely held against the palate.

Jaw movement for speech is rather different than jaw movement for other functions, such as chewing or swallowing. Researchers see somewhat different patterns in the movement of the jaw if a subject reads or eats, with speech showing less rhythmic, low-amplitude movements.

The jaw consists of a large curved bone with two perpendicular processes (called *rami*, or branches) that rise up to meet the skull. The lower section contains the chin (or *mental protuberance*) and holds the teeth. Figure 4-7 shows the anatomical view of a jaw.

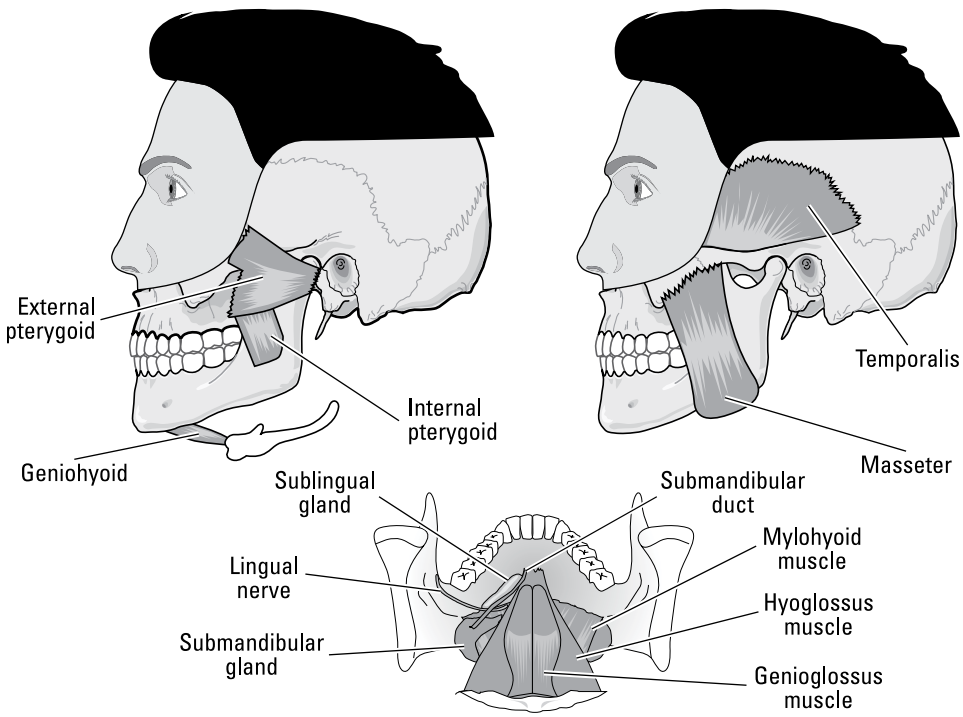


Figure 4-7: A jaw and its muscles.

The rami meet the skull at the *temporomandibular joint (TMJ)*. The jaw has two TMJ (one on each side of the skull) that work in unison. These complex joints allow a hinge-like motion, a sliding motion, and a sideways motion of the jaw. You may have heard of TMJ because of *TMJ disorder*, a condition in which the TMJ joint can be painful and audibly pop or click during certain movements.

A series of muscles, known as the *muscles of mastication,* move the jaw. These muscles include the *masseter, temporalis, and internal pterygoid* (all of which raise the jaw), and the *external pterygoid, anterior belly of digastric* (not shown in the figure)*, mylohyoid,* and *geniohoid* (all of which lower the jaw). Look at Figure 4-7 to see these muscles.

Eyeing the soft palate and uvula: The velum

You find the velum, which consists of the soft palate and uvula, behind your hard palate (see Figure 4-1). *Velum* means *curtain* and is a hanging flap in the back of the roof of the mouth. The soft palate is called "soft" because it has cartilage underlying it, instead of bone, and the *uvula* (the structure at the back of the velum that hangs down in the throat; refer to the next section for more details). You can feel this difference if you probe this part of your palate with your tongue. The uvula is a structure used for consonant articulations (such as trills) in some languages.

The velum is an important place of articulation for many English speech sounds, including /k/, /g/, /ŋ/, /ɬ/, and /w/, as in the words "*k*ick," "*g*host," "ri*ng*," "pi*ll*," and "*w*et".

Like the tongue, the velum is highly coordinated and capable of quick and fine-tuned movements. An important velar function is to open and close the *velopharyngeal port* (also known as the *nasal port*), the airway passage to the nasal cavity. This function is necessary because most speech sounds are non-nasal, so it's important that most air not flow out the nose during speech.

Both passive and active forces move the velum:

- **Passive:** The velum is acted on by gravity and airflow.
- **Active:** A series of five muscles move the velum in different directions. The five muscles are *palatal levator, palatal tenser, uvulus, glossopalatine,* and *pharyngopalatine.*

The path of the velum moving up and down during speech is fascinating to watch (look at `www.utdallas.edu/~wkatz/PFD/phon_movies.html`). The moving velum has a hooked shape with a dimple in the bottom as it lifts to close the nasal port. Every time you make a non-nasal oral sound, you

subconsciously move your velum in this way. When a nasal is made, however, as in /ɑnɑ/, your velum moves forward and down, allowing air passage into the nasal cavity.

The velum actually doesn't act alone. Typically, the sides and the back wall of the *pharnyx* (the back of your throat) participate with the closure to form a flap-like sphincter motion. Different people seem to make this closure in slightly different ways.

Going for the grapes: The uvula

The uvula (which means "bunch of grapes") hangs down in the back of the throat. It's that part that cartoonists love to draw! This region of the velum has a rather rich blood supply, leading anatomists to suspect that it may have some cooling function. In terms of speech, some languages use this part of the body to make trills or fricatives (flip to Chapter 16 for additional information). However, English doesn't have uvular sounds.

Pondering Speech Production with Models

Ordinary conversational speech involves relaying about 12 to 18 meaningful bits of sound (technically referred to as *phonemes*) per second. In fast speech, this rate is easily doubled. Such rates are much faster than anyone can type on a keyboard or tap out on a cell phone.

In order for you to produce speech, your mind sends ideas to your mouth at lightning speed. According to Professor Joseph Perkell of MIT, approximately 50 muscles governing vocal tract movement are typically coordinated to permit speaking, so that you can be understood. And this estimate of 50 muscles, by the way, doesn't even include the muscles of the respiratory system that are also involved.

You must coordinate all these muscles for speech without requiring too much effort or concentration so that you can complete other everyday tasks, such as tracking your conversation, walking around, and so on.

Being able to understand healthy speech production is important so that clinicians can better assist individuals with disordered or delayed speech processes. To grasp how people can accomplish this feat of talking, scientists make observations and build models. The following sections examine some of these different models.

Models are essential to science

A *model* is a visual representation, whether physical or mathematical, that helps scientists study something in more depth. In particular, it allows scientists to test *hypotheses* about *theories.* A *theory* is a general set of underlying principles and assumptions concerning the natural world that has arisen from repeated observations and testing. A *hypothesis* is a specific, testable prediction about what you would expect to happen, given a certain theory.

For example, a phonetician wishes to test the hypothesis that children boost vowel intelligibility by varying their fundamental frequency to a greater extent than adults. This hypothesis follows from the source-filter theory of speech production. Phoneticians can generate a statistical model of the vocal tract and compare findings for children and adults.

Ordering sounds, from mind to mouth

Speech is the predominant channel people use to relay language. Other channels include reading/writing, and sign language. Because speech sounds don't hang around for anyone to see like written communication, the order in which sounds are produced is critical.

Speech sounds aren't strung together like beads on a string; the planned sounds blend and interweave by the time they reach the final output stage by a principle called *coarticulation.* Two main types of coarticulation, which are as follows, affect sound production:

- **Anticipatory:** Also referred to as *look-ahead* or *right-to-left coarticulation,* it measures how a talker prepares for an upcoming sound during the production of a current sound. It's considered a measure of speech planning and shows many language-specific properties.
- **Perseverative:** Also referred to as *carry-over* or *left-to-right coarticulation,* it describes the effects of a previously made sound that continue onto the present sound. Think of a nagging mother-in-law who is still sticking around when she shouldn't be there any more. *Perseverative coarticulation* measures the physical properties of the articulators, or in other words how quickly they can be set to move or stop after being set into motion. For example, if you say "I said *he* again," the breathiness of the /h/ will carry over into the vowel /i/. Such breathiness doesn't carry over from a preceding sound that isn't breathy, such as the /b/ in the word "*b*ee" (/bi/ in IPA).

All people coarticulate naturally while they speak, in both anticipatory and perseverative directions. Refer to Chapter 6 for more on coarticulation.

Speech is also redundant, meaning that information is relayed based on more than one type of clue. For example, when you make the consonant /p/ in the

word "*p*et," you're letting the listener know it's a /p/ (and not a /b/) by encoding many types of acoustic clues, based on frequency and timing (refer to Chapters 15 and 16 for more specifics). In this way, humans are quite different than computers. Humans usually include many types of information in speech and language codes before letting a listener get the idea that a distinction has been made.

Controlling degrees of freedom

To understand how speech is produced, researchers have long tried to build speech systems and have often been humbled by the ways in which these approaches have come up lacking. The *degrees of freedom* problem, which is that many muscles fire in a complex order to produce speech, is so difficult that scientists have tried to make some sense of it.

Because speech science researchers have known for quite some time about basic speech anatomy, they have searched for muscle-by-muscle coordination of speech. Scientists first hoped that by studying a single muscle (or small group of muscles) they could explain in a simple fashion how speech was organized. Electrodes were available for recording muscle activity, and scientists hoped that by charting the time course of muscle activation, they could get a better idea of how speech was planned and regulated. For instance, they searched for the pulse trains involved in stimulating the *orbicularis oris,* the facial muscles, the respiratory muscles, the intrinsic lingual muscles, the extrinsic lingual muscles, and so on in a certain order. They presumed that the brain's neural structures coordinated all the steps.

However, the data instead suggested that speech is much more complex. There are too many processes for the brain to regulate centrally, and the brain doesn't trigger muscles in a sequential, one-by-one fashion.

This degrees of freedom problem is ongoing in speech science. For this reason, scientists have abandoned the view that individual muscle actions are programmed in running speech on a one-by-one basis. Instead, researchers have taken other steps, building models that are organized more functionally, along coordinative structures or gestures. Researchers have tried to re-create how these processes happen, either in a mathematical model, in a graphic simulation (such as an avatar), in a mechanical robot, or in a computerized neural model.

In models, scientists describe trade-offs between sets of muscles to achieve a common function such as lip closure. These muscles are hierarchically related such that a speech-planning mechanism only need trigger a function such as *elevate lip*, which would trigger a whole complex of muscles in the face, lips, and jaw. Scientists have found much evidence for this type of *synergistic* (working together for an enhanced effect) model. For instance, lip-closing muscles do work in synergy with the muscles of the face and jaw; if some muscles are interrupted in function, others take over. Thinking the body has some type of central executive that needs to plan each muscle's activity (on an individual basis) just doesn't make sense.

Trying to map speech

At some level, researchers hope to connect the lofty world of language (say, thinking of the lines of a Shakespeare play) and actually *saying* some of these words in the messy reality of speech, motor control, acoustics, and perception. This problem clearly isn't easy to solve. If it were, society would already have convincing talking robots or computers without keyboards that people could chat with like any other person.

In early work on this problem, linguists assumed that the same divisions used to describe language (words, syllables, morphemes, consonants and vowels, and features) naturally mapped to speech goals. As a starting point, researchers thought the process was basically linear, from start to finish. According to this view, speech would be accomplished with a left to right readout, having a short-term buffer that allowed for syllables.

Mounting evidence shows that speech isn't produced in such a linear fashion and that linguistic concepts aren't generally adequate for describing the complexity of speaking. For instance, the /dʒ/ in "Jerry" is realized by placing the tongue against the alveolar ridge and releasing into a post-alveolar fricative while voicing. This physical action requires dozens of muscle sets in the vocal tract, plus respiratory muscles. Somehow, labeling this sound as another linguistic feature doesn't seem to satisfy many researchers that the process is really being explained.

Feeding forward, feeding back

Scientists assume that people speak by mapping information from higher to lower processing levels, which is called *feed-forward processing*. You start with a concept, find the word (*lexical selection*), map the word into its speech sounds (phonemes), and finally output a string of spoken speech. In feed-forward processing, information flows without needing to loop back. In terms of speech production, feed-forward mechanisms include your knowledge of English, your years of practice speaking and moving your articulators, and the automatic processes used to produce speech. This overlearned aspect of speech makes its production effortless under ordinary conditions. Feed-forward processing is rapid because it doesn't require a time delay such as feed-back processing.

However, you also need feedback processes; you don't talk in a vacuum. You hear yourself talk and use this information to adjust your volume and rate. You also sense the position of lips, tongue, jaw, and velum. You, along with nearly everyone else, use this type of feedback to adjust your ongoing speech.

People can rely on auditory feedback to make adjustments. For example, if you're at a party where the background sound is loud, you'll probably start speaking louder automatically. If suddenly the sound drops, you can lower your volume. You also hear the sound of your voice through the bones of your skull, which is called *bone conduction*. For this reason, when you hear your voice audio-recorded, you sound different, often tinnier.

In terms of articulatory feedback, a visit to the dentist can provide some insight. Numbing the tongue with anesthetic reduces articulatory feedback and compromises the production of certain sounds.

A good way to visualize the process is to imagine a house thermostat. A simple, old-fashioned version will wait until your room gets too cold in the winter before kicking on the heat. When the room gets too hot, the thermostat kicks it off. This is feedback — accurate, but time consuming, clunky (and not really smart). Some people have smarter thermostats that incorporate feed-forward information. You can set such a thermostat, for example to turn down the heat when you're away during the day or asleep at night (ahead of time) and then adjust it back to comfortable levels when you're home or active again.

Coming Up with Solutions and Explanations

Understanding speech production is one of the great scientific challenges of this century. Scientists are using a variety of approaches to understand how speech is produced, including systems that allow for precise timing of speech gestures and computational models that incorporate brain bases for speech production. This section gives you a taste of these recent approaches.

Keeping a gestural score

Figuring out how speech can be controlled is important, but it still doesn't solve the problem of degrees of freedom, or basically how 50 or so odd sets of muscles coordinate during fluent speech.

In 1986, researchers at Haskins Laboratory proposed to track speech according to a gestural score, which other researchers have modeled. With a *gestural score,* for a word in the mind to be finally realized as speech, you begin with a series of articulatory gestures. They include adjustments to your speech anatomy such as lip protrusion, velar lowering, tongue tip and body positioning, and adjustment of glottal width. Each gesture is then considered a sequence within an articulatory score (much like different measures might be thought to be parts of a musical composition). However, in this model the articulatory gestures have time frames expressed as *sliding windows* within which the gestures are expressed. By lining up the sliding windows of the various articulatory gestures over time, one can read out an action score for the articulation of a spoken word.

You can find more information on gestural scores, including an example for the word "pan" at `www.haskins.yale.edu/research/gestural.html`.

This type of model can capture the graded, articulatory properties of speech. Scientists can combine such models with linguistic explanations and computer and anatomical models of speech production.

Connecting with a DIVA

Frank Guenther, a professor at Boston University, developed the Directions Into Velocities of Articulators (DIVA) model to study speech production. DIVA incorporates auditory and *somatosensory* (body-based) feedback in a distributed neural network.

Neural network models are very basic simulations of the brain, set up in computers. A *neural network* consists of many artificial neurons, each of which gets stimulated and fires (electronically), acting in the computer as if it were somehow a human neuron. These neurons are linked together in nets that feed their information to each other. For instance, in a feed-forward network, neurons in one layer feed their output forward to a next layer until one gets a final output from the neural network. In many systems an intermediate layer (called a *hidden layer*) helps process the input and output layers.

These nets are capable of some surprising properties. For instance, they can be shown a pattern (called a *training set*) and undergo supervised learning that will eventually allow them to complete complex tasks, such as speech production and perception.

Components of the DIVA model are based on brain-imaging data from studies of children and adults producing speech and language, thus relating speech-processing activity with what scientists know about the brain. DIVA *learns* to control a vocal tract model and then sends this information to a speech synthesizer. Researchers can also use DIVA to simulate MRI images of brain activation during speech, against which the patterns of real talkers can be compared.

The first DIVA models were only able to simulate single speech sounds, one by one. However, a more recent model, called gradient order DIVA (GODIVA) can capture sequences of sounds. As models of this type are elaborated, they may offer new insights into how healthy and disordered people produce and control speech sounds.

Chapter 5

Classifying Speech Sounds: Your Gateway to Phonology

In This Chapter

- Taking a closer look at features
- Noting odd things with markedness
- Keying in on consonant and vowel classification
- Grasping the important concepts of phonemes and allophones

Naming is knowledge. If you classify a speech sound, you know what its voicing source is, where it is produced in the vocal tract, and how the sound was physically made. This chapter introduces you to how speech sounds are described in phonetics. I discuss some of the traditional ways that phoneticians use to classify vowels and consonants — ways that are used somewhat differently across these two sound classes. I dedicate a major part of this chapter to the concepts of phoneme and allophone, important building blocks needed to understand the *phonology* (sound systems and rules) of any language.

Focusing on Features

A *phonetic feature* is a property used to define classes of sounds. More specifically, a feature is the smallest part of sound that can affect meaning in a language. In early work on feature theory, phoneticians defined features as the smallest units that people listened to when telling meaningful words apart, such as "dog" versus "bog." As work in this area progressed, phoneticians also defined features by the role they played in *phonological rules,* which are broader sound patterns in language (refer to Chapters 8 and 9 for more on these rules). The following sections discuss the four types of phonetic features.

Binary: You're in or out!

You may be familiar with the term *binary* from computers, meaning having two values, 0 or 1. Think of flipping a light switch either on or off. Because binary values are so (blessedly) straightforward, engineers and logicians all over the world love them. Phonologists use binary features because of their simplicity and because they can be easily used in computers and telephone and communication systems.

An example of a binary feature is *voicing*. A sound is either *voiced* (coded as + in binary features) or *voiceless* (coded as –). Another example is *aspiration*, whether a stop consonant is produced with a puff of air after its release. Using binary features, phoneticians classify stop consonants as being "+/– aspiration."

To see how binary features are typically used for consonants and vowels, Figure 5-1 shows a binary feature matrix for the sounds in the word "needs," written in IPA as /nidz/.

	/n/	/i/	/d/	/z/
Syllabic	–	+	–	–
Consonantal	+	–	+	+
High	–	+	–	–
Back	–	–	–	–
Low	–	–	–	–
Anterior	+	–	+	+
Coronal	+	–	+	+
Round	n/a	–	n/a	n/a
Tense	n/a	+	n/a	n/a
Voice	+	(+)	+	+
Continuant	+	(+)	–	+
Nasal	+	(–)	–	–
Strident	–	n/a	–	+
Lateral	–	n/a	–	–

Figure 5-1: The word "needs" represented in a binary feature matrix.

Illustration by Wiley, Composition Services Graphics

In this figure, the sound features of each phoneme (/n/, /i/, /d/, and /z/) are listed as binary (+/–) values of features, detailed in the left-most column. For example, /n/ is a consonant (+ consonantal) that doesn't make up the nucleus of a syllable (– syllabic). The next three features refer to positions of the tongue body relative to a neutral position, such as in production of the vowel /ə/ for "the". The consonant /n/ is negative for these three features. Because /n/ is produced at the alveolar ridge, it's considered + *anterior* and + *coronal* (sounds made with tongue tip or blade). Because /n/ isn't a vowel, the features "round" and "tense" don't apply. /n/ is produced with an ongoing flow of air and is thus + *continuant*. It's + *nasal* (produced with airflow in the nasal passage), not made with noisy hissiness (– strident) nor with airflow around the sides of the tongue (– lateral).

If you're an engineer, you can immediately see the usefulness of this kind of information. Binary features, which are necessary for many kinds of speech and communication technologies, break the speech signal into the smallest bits of information needed, and then discard and eliminate the less useful information.

Phoneticians only want to work with the most needed features. For instance, because most stop consonants are *oral stops* (sounds made by blocking airflow in the mouth, refer to Chapter 6 for more information), you don't usually need to state the oral features for /p/, /t/, /k/, /b/, /d/, and /g/. However, the *nasal feature* (describing sounds made with airflow through the nasal passage) is added to the description of the (less common) English nasal stop consonants /m/, /n/, and /ŋ/. Here are some examples of reducing this feature *redundancy* (repetition) to make phonetic description more streamlined and complete:

> /b/**:** This sound is typically described as a voiced bilabial stop. You don't need to further specify "oral" because it's understood by default.
>
> /m/**:** This sound is typically described as a *voiced bilabial nasal* or a *voiced bilabial nasal stop*. Because nasals are less common sounds and are distinguished from the more typical oral stops by their nasality, it's important to note "nasal" in their description.

Here is another example. In Figure 5-1, the last 5 features (voice, continuant, nasal, strident, lateral) apply chiefly to consonants. Thus, the vowel /i/ (as in "eat"), doesn't need to be marked with these (+ voice, + continuant, and so on). For this reason, I've placed the values in parentheses or marked them as "n/a" (not applicable).

Graded: All levels can apply

Other properties of spoken language don't divide up as neatly as the cases of voicing and aspiration, as the previous section shows. Phoneticians typically use *graded* (categorized) representations for showing various melodic patterns across different intended meanings or emotions. *Suprasegmental* (larger than the individual sound segment) properties (such as stress, length, and intonation) indicate gradual change over the course of an utterance.

For example, try saying "Oh, really?" several times, first in a surprised, then in a bored voice. You probably produced rather different melodic patterns across the two intended emotions. Marking these changes with any kind of simple binary feature would be difficult. That's why using graded representations is better. Here is this graded example:

Surprised: "Oh really"

Bored: "Oh really"

To represent the melody of these utterances, you have a couple of different options. You can draw one of the following two:

- **Pitch contour:** A *pitch contour* is a line that represents the fundamental frequency of the utterance. Figure 5-2 provides an example.

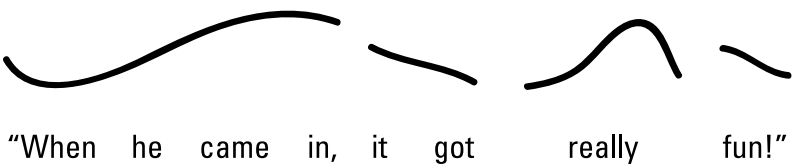


Figure 5-2: A pitch contour example.

Illustration by Wiley, Composition Services Graphics

- **Numeric categorization scheme:** In such a representation (as this), numeric levels of pitch (where 1 is low, 2 is mid, and 3 is high) and the spacing between numbers representing *juncture* (the space between words) provide a graded representation of the information.

```
                  1   3   2
Surprised:        "Oh really"

                  1         1   1
Bored:            "Oh       really"
```

There is no one correct method for transcribing suprasegmental information described in the IPA. However, refer to Chapters 10 and 11 for some recommendations.

Articulatory: What your body does

Articulatory features refer to the positions of the moving speech *articulators* (the tongue, lips, jaw, and velum). In the old days, articulatory features also referred to the muscular settings of the vocal tract (tense and lax). The old phoneticians got a lot right; the positions of the speech articulators are a pretty good way of classifying consonant sounds. However, this muscular setting hypothesis for vowels was wrong. Phoneticians now know the following:

- **For consonant sounds:** Articulatory features can point to the tongue itself, such as *apical* (made by the tip), *coronal* (made by the blade), as well as the regions on the lips, teeth, and vocal tract where consonantal constrictions take place (bilabial, labiodental, dental, alveolar, post-alveolar, retroflex, palatal, uvular, pharyngeal, and glottal).
- **For vowels:** Articulatory descriptions of vowels consider the height and backness of the tongue. Tongue position refers to high, mid, or low (also known as having the mouth move from close to open) and back, central, or front in the horizontal direction. Figure 5-3 shows this common expression in a diagram known as a vowel chart, or vowel quadrilateral.

  Vowel charts also account for the articulatory feature of *rounding* (lip protrusion), listing unrounded and rounded versions of vowels side by side. For instance, the high front rounded vowel /y/, as found in the French word "tu" (meaning *you* informal), would appear next to the high front vowel /i/. English doesn't have rounded and unrounded vowel pairs. Instead, the four vowels with some lip rounding are circled in Figure 5-3. The arrows show movement for *diphthongs* (vowels with more than one quality). Chapter 7 provides further information about vowels, diphthongs, and the vowel quadrilateral.

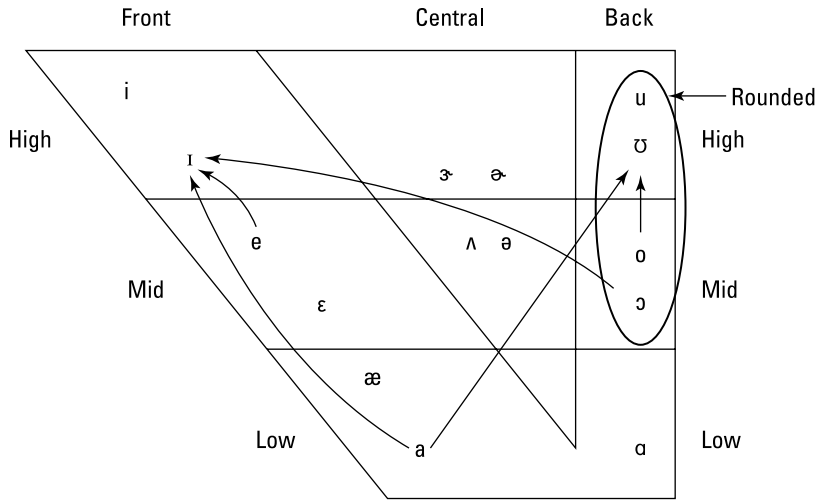


Figure 5-3: Vowel quadrilateral showing rounded vowels and diphthongs.

Illustration by Wiley, Composition Services Graphics

Acoustic: The sounds themselves

Although specifying more or less where the tongue is during vowel production is okay for a basic classification of vowels, doing so doesn't cover everything. Phoneticians agree that *acoustic* (sound-based) features give a more precise definition, especially for vowels. These acoustic features have to do with specific issues, such as how high or low the frequencies of the sounds are in different parts of the sound spectrum, and the duration (length) of the sounds.

Vowels in the past: Getting tense about lax

Phoneticians used to think that vowels called *tense* were produced with more muscular tension than the vowels called *lax*. In the 1960s, instrumentation became available for investigating muscle activation during speech, and experiments in the 1970s checked out the tense versus lax question. By the way, some of these experiments were probably not a whole lot of fun for the subjects because "hook-wire" electrodes were used (fish-hook like electrodes that are injected into the muscles of the tongue, cheeks, and throat). Ouch!

The results provided no evidence that the tense vowels are produced with any more muscle activation than lax vowels. Today, phoneticians consider the English vowel tense/lax difference to be a phonological one. The English tense vowels are those that can be produced in stressed *open syllables,* that is, without any consonant at the end. Thus, you can say "bee" (/bi/ in IPA) or "shoe" (/ʃu/ in IPA), but you can't say /bɪ/ or /ʃʊ/ and have them be English words. People can use lax vowels in *closed syllables,* syllables ending with a consonant (such as "bit" /bɪt/ and "shook" /ʃʊk/).

The tongue makes many different shapes when you say vowels, and a more critical determining factor in what creates a vowel sound is the shape of the tube in your throat. Refer to the top part of Figure 5-4 for a sample of these tube shapes.

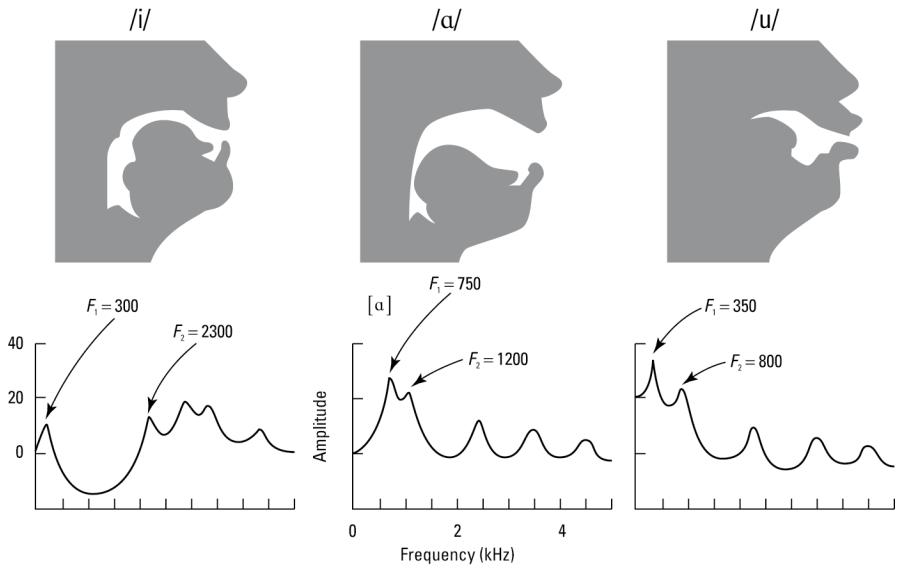


Figure 5-4: Three cross-sectional heads showing different tube shapes and the corresponding vowel spectra.

Illustration by Wiley, Composition Services Graphics

To work with acoustic features, phoneticians analyze speech by computer and look for landmarks. One such important landmark for vowels is called *formant frequencies,* which are peaks in the spectrum which determine vowel sound quality. Chapter 12 explains more about acoustic features and formant frequencies.

Marking Strange Sounds

The number of possible features for any given speech sound can become, well, many! As a phonetician considers the numerous sounds in language, it becomes important to keep track of which are the more common sounds, those likely to be universal across the world's languages, and which sounds are rare — that is, the oddballs of the phonetics world.

To do so, the unusual sound or process is considered *marked,* whereas the rather common one is *unmarked.* Here are some examples:

- Stop consonants made at the lips (such as /p/ and /b/) are relatively common across the world's languages, and are thus rather unmarked. However, the first sound in the Japanese word "Fuji" is a voiceless bilabial fricative made by blowing air sharply through the two lips. This fricative (classified with the Greek character "*phi*," /ɸ/ in IPA) is relatively rare in the world's languages, and is thus considered marked.
- The vowels /i/, /u/, and /a/ are highly unmarked, because they're some of the most likely vowels to be found in any languages in the world. In contrast, the rounded vowels /y/, /ø/, and /œ/ are more marked, because they only tend to appear if a language also has a corresponding unrounded series /i/, /e/, and /a/.

How a phonetician determines whether a sound is marked or unmarked is a pretty sophisticated way of viewing language. Saying that a sound or process is marked means that it's less commonly distributed among the world's languages, perhaps because a certain sound is relatively difficult to hear or is effortful to produce (or both).

However, remember that a phonetician talking about markedness is quite different than people saying that a certain language is difficult. The idea of a language being difficult is usually a value judgment: It depends on where you're coming from. When deciding whether a language is simple or complex, be careful about making value judgments about other languages. For example, Japanese may seem like a "difficult" language for an English speaker, but perhaps not so much for a native speaker of Korean because Japanese and Korean share many phonological, syntactic, and writing similarities that English doesn't share.

Also, before making a judgment of difficulty, think about what part of the language is supposed to be difficult. Linguists talk about languages in terms of their phonology, *morphology* (way of representing chunks of meaning), *syntax* (way of marking who did what to whom), *semantics* (phrase and sentence level meaning), and writing systems, assuming the language has a written form (most languages in the world don't have a written form). It's very typical for languages to be complex in some areas and not in others. For instance, Japanese has a rather simple sound inventory, a relatively straightforward syntax, but a very complicated writing system. In contrast, Turkish has a fairly simple writing system but a rather complex phonology and syntax.

Introducing the Big Three

In order to grasp a basic tenet of phonetics, you need to know about the Big Three — the three types of articulatory features that allow you to classify consonants. For phonetics, the three are voicing, place, and manner, which create the acronym VPM. Here is a bit more about these three and what you need to know:

- **Voicing:** This term refers to whether or not the vocal folds are buzzing during speech. If there is voicing, buzzing occurs and speech is heard as voiced, such as the consonants in "*b*ee" (/bi/) and "*z*oo" (/zu/). If there is no buzzing, a sound is voiceless, such as the consonants in "*p*i*t*" (/pɪt/) or "*sh*y" (/ʃaɪ/). All vowels and about half of the consonants are normally produced voiced, unless you're whispering.
- **Places of articulation:** This term relates to the location of consonant production. They're the regions of the vocal tract where consonant constriction takes place. Refer to Table 5-1 for the different places.

Table 5-1 Where English Consonants Are Produced

Feature	*Location*	*IPA*
Bilabial	At the two lips	/p/, /b/, /m/
Labiodental	Lower lip to teeth	/f/, /v/
Dental	Teeth	/θ/, /ð/
Alveolar	Ridge on palate behind teeth	/s/, /z/, /t/, /d/, /ɹ/, /l/, /n/
Post-alveolar (also known as palato-alveolar)	Behind the alveolar ridge	/tʃ/, /dʒ/, /ʃ/, /ʒ/
Palatal	At the hard palate	/j/
Velar	At the soft palate	/k/, /g/, /ŋ/
Labio-velar	With lips and soft palate	/w/
Glottal	Space between vocal folds	/ʔ/, /h/

- **Manner of Articulation**: This term refers to the how of consonant production, specifically, the nature of the consonantal constriction. Table 5-2 lists the major manner types for English.

Table 5-2 How English Consonants Are Produced

Name	*Construction Type*	*IPA*
Stop	Complete blockage – by default, oral	/p/, /t/, /k/, /b/, /d/, /g/,
Nasal	Nasal stop – oral cavity stopped, air flows out nasal cavity	/m/, /n/, /ŋ/
Fricative	Groove or narrow slit to produce hissing	/θ/, /ð/, /ʃ/, /ʒ/, /s/, /z/, /h/, /f/, /v/

(continued)

Table 5-2 *(continued)*

Name	*Construction Type*	*IPA*
Affricate	Combo of stop and fricative	/tʃ/, /dʒ/
Approximant	Articulators approximate each other, come together for a "wa-wa" effect	/w/, /ɹ/, /l/, /j/
Tap	Brief complete blockage	/ɾ/
Glottal stop	Complete blockage at the glottal source	/ʔ/

Every time you encounter a consonant, think of VPM and be prepared to determine its voicing, place, and manner features.

Making flashcards is a great way to master consonants and vowels, with a word or sound on one side, and the features on the other.

Moving to the Middle, Moving to the Sides

Most speech sounds are made with *central airflow,* through the middle of the oral cavity, which is the default or unmarked case. However for some sounds, like the "l" sound, a *lateral* (sideways) *airflow* mechanism is used, which involves air flowing around the sides of the tongue.

In English, you can find an important central versus lateral distinction for the voiced alveolar approximants /ɹ/ and /l/. You can hear these two sounds in the minimal pair "*l*eap" and "*r*eap" (/lip/and /ɹip/). For /l/, air is produced with lateral movement around the tongue.

To test it, try the phonetician's cool air trick. To use this test, produce a speech sound you wish to investigate, freeze the position, and suck in air. Your articulators can sense the cool incoming air, and you should be able to get a better sensation of where your tongue, lips, and jaw are during the production of the sound. For this example, to do this test, follow along:

1. **Say "reap," holding the initial consonant (/ɹ/).**
2. **Suck in some cool air to help feel where your tongue is and where the air flows.**
3. **Say "leap," doing the same thing while sensing tongue position and airflow for the initial /l/.**

You should be able to feel airflow around the sides of the tongue for /l/. You may also notice a bit of a duck-like, slurpy quality to the air as it flows around the sides of the tongue. This is a well-known quality, also found as a feature in some of the languages that have slightly different lateral sounds than are found in English. Chapter 16 provides more information on these unusual lateral sounds.

Sounding Out Vowels and Keeping Things Cardinal

Knowing what phoneticians generally think about when classifying vowels is important. In fact, phonetics has a strong tradition, dating back to 19th century British phonetician Daniel Jones, of using the ear to determine vowel quality. An important technique for relying on the ear depends on using *cardinal vowels,* vowels produced at well-defined positions in articulatory space and used as a reference against which other vowels can be heard.

Figure 5-5 shows how cardinal vowels work. Plotted are the cardinal vowels, as originally defined by Jones and still used by many phoneticians today. These vowels aren't necessarily the vowels of any given language, although many lie close to vowels found in many languages (for instance, cardinal vowel /i/ is quite close to the high front unrounded vowel of German). The relative tongue position for each vowel is shown on the sides of the figure.

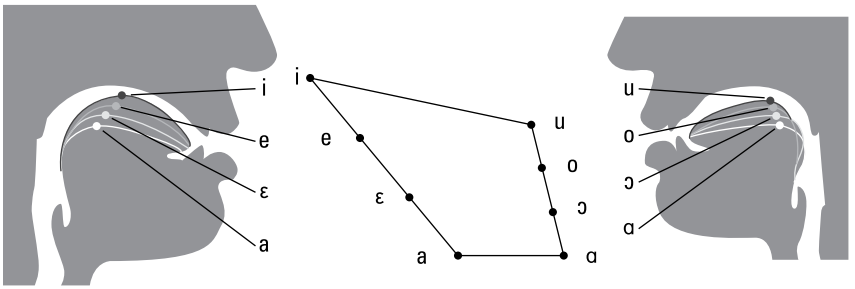


Figure 5-5: English cardinal vowels and associated tongue positions.

Illustration by Wiley, Composition Services Graphics

To make cardinal vowel /i/, make a regular English /i/ and then push your tongue higher and more front — that is, make the most extreme /i/ possible for you to make. This *point vowel,* or extreme articulatory case, is a very pure /i/ against which other types of "/i/-like" vowels may be judged. With such an extremely /i/-sounding reference handy, a phonetician can describe how the high front sounds of, say, English, Swedish, and Japanese differ.

The same type of logic holds true for the other vowels in this figure, such as the low front vowel /a/ or the high back vowel /u/. Just like with the regular IPA chart (see Chapter 3), this set of cardinal vowels also has *rounded* and *unrounded* (either produced with the lips protruded or not) vowels. Jones called the rounded series the *secondary cardinal vowels.*

To hear Daniel Jones producing 18 cardinal vowels (from an original 1956 Linguaphone recording), go to `www.youtube.com/watch?v=haJm2QoRNKo`.

Tackling Phonemes

A *phoneme* is the smallest unit of sound that contributes to a meaning in a language. Knowing about phonemes is important and frequently overlooked by beginning students of phonetics because they can seem so obvious and, well, boring. However, phonemes aren't boring. In fact, they're essential to many fields, such as speech language pathology, psycholinguistics, and child language acquisition.

In simple terms, a *phoneme* is psychological. If you want to talk about a speech sound in general, it's a *phone*, not a phoneme. A sound becomes a phoneme when it's considered a meaningful sound in a language. Phoneticians talk about phonemes of English or Russian or Tagalog. That is, to be a phoneme means to be a crucial part of a particular language, not language in general.

Furthermore, one person's phoneme isn't necessarily another person's phoneme. If I were to suddenly drop you among speakers of a very different-sounding language, and these people tried to teach you their language's sound system, you would probably have a difficult time telling certain sounds apart. This is because the sound boundaries in your mind (based on the phonemes of your native language) wouldn't work well for the new language I have dumped you in.

If you're a native English speaker, you'd be in this plight if you were trying to hear the sound of the Thai consonant /t/ at the beginning of a syllable. For example, the clear spicy Thai soup "*t*om yum" may sound to you as if it were pronounced "*d*om yum," instead of having an unaspirated /t/ at the beginning. Native Thai speakers may be surprised and even amused at your inability to hear this word pronounced correctly.

Determining whether speech breaks down at the phonemic level is important in understanding language disorders such as *aphasia*, the language loss in adults after brain damage, and in studying child language acquisition. The following sections take a closer look at phonemes.

Defining phonemes

To investigate the sound system of a language, you search for a phoneme. To be a phoneme, a sound must pass two tests:

- **It must be able to form a minimal pair.** A *minimal pair* is formed whenever two words differ by one sound, such as "bat" versus "bag" (/bæt/ and /bæg/), or "eat" versus "it" (/it/ and/ɪt/). In the first pair, consonant voicing (/t/ versus /g/) makes the difference. In the second pair, vowel quality (/i/ versus /ɪ/) makes the difference. However, in both cases a single phoneme causes a meaningful distinction between two words. Phoneticians consider minimal pairs a test for a distinctive feature because the feature contributes to an important, sound-based meaning in a language.
- **It should be in free (or contrastive) distribution.** The term *free distribution* means a sound can be found in the same environment with a change in meaning. For example, the minimal pair "bay" versus "pay" (/be/ and /pe/) show that English /b/ and /p/ are in free distribution.

  Notice that phonemes in a language (such as the English consonants /s/, /t/, /g/, and the vowels /i/, /ɑ/, and /u/) can appear basically anywhere in a word and change meaning in pretty much the same fashion. The same kind of sound-meaning relationships hold true even when these sounds are in different syllabic positions, such as "*t*oe" versus "*g*o" (initial position) or "sea*t*" versus "see*d*" (final position).

Complementary distribution: Eyeing allophones

Complementary distribution is when sounds don't distribute freely, but seem to vary systematically (suggesting some kind of interesting, underlying reason). Complementary distribution is the opposite of free distribution, a property of phonemes. The systematically varying sounds that result from complimentary distribution are called *allophones,* a group of possible stand-ins for a phoneme. It's kind of like Clark Kent and Superman — they're really the same guy, but the two are never seen in the same place together. One can stand in for the other.

The prefix *allo-* means a systematic variant of something, and *-phone* is a language sound. Therefore, an allophone is a systematic variant of a phoneme in language. In this case, a language has one phoneme of something (such as a "t" in English), but this phoneme is realized in several different ways, depending on the context.

Sticking to the rules of phonology

Part of speaking a language is internalizing its *phonology*, the systematic sound rules. A speaker of American English would know the rules that govern which "t" to use and when, and would be able to use them automatically. When faced with a new (made-up) word, such as "*t*elps," she would pronounce the initial "t" with aspiration, whereas she wouldn't do the same to the "t" at the end of "kra*t*".

Put your hand under your mouth and try for yourself. You should feel a puff of air on the first "t" of "*t*elps" and none on the "t" of "kra*t*". You probably used two different allophones of the phoneme /t/ (that is, [t^h] and [t]) because you know General American English.

Phonologists love figuring out which sounds in a language *corpus* (body or sample of a language) represent *phonemes* (meaningful sounds of the language) and which are allophones of a single, underlying phoneme. To search for phonemes, people look for minimal pairs and free or contrastive distribution. To search for allophones, phonologists hunt for sounds that are similar phonetically (for example, like /s/ and /ʃ/ or /t/ and /tʃ/) and which also show complementary distribution.

English has just one meaningful "t". At the level of meaning, the "t" in "*T*ed" is the same as the "t" in "ba*t*," in "Be*tt*y," and in "mi*tt*en." They all represent some kind of basic "t" in your mind. However, what may surprise you is that each of the "t" sounds for these four words is pronounced quite differently, as in the following:

Word	***IPA Transcription (narrow)***	***The "t" Used (Allophone)***
Ted	[t^hɛd]	aspirated t
bat	[bæt]	unaspirated t
Betty	[ˈbɛɾɪ]	alveolar flap
mitten	[ˈmɪʔn̩]	glottal stop

Each of these words only has one meaningful "t" sound, but depending on the context, each word has its own realized but different kind of "t" sound.

To put it another way, you understand just one phoneme /t/, but actually speak and hear four different *allophones*. These include aspirated t, unaspirated t, alveolar flap, and glottal stop. Each of these allophones is a systematic variant of the phoneme /t/ in General American English. ***Note:*** Although phonemes are written in slash brackets (/t/), allophones are written in square brackets, ([t]).

Sleuthing Some Test Cases

Making sure you have the concepts of phoneme and allophone is important and one way to do so is to examine other languages. In these sections, I conduct a brief phonological analysis of English and contrast these patterns with those of with Spanish and Thai. I also provide an American indigenous language example.

Comparing English with Thai and Spanish

Here I make a quick comparison of how two other languages treat their stop consonants, in comparison to English. Table 5-3 focuses on the voiceless, bilabial stop (/p/ in IPA) and compares English with Thai and Spanish examples.

Table 5-3 Phonemes and Allophones for "p" in English, Spanish, and Thai

Language		*IPA Symbols*	*Examples*	
English	One phoneme, two allophones	/p/ --> [p^h] or [p]	[p^hɛt] "pet"	[næp] "nap"
Thai	Two phonemes	/p/, /p^h/	[p^ha:] "forest"	[pa:] "split"
Spanish	One phoneme	/p/		[ˈpɛɾo] "but"

The English /p/ has an aspirated form found at the beginning of syllables (such as "*p*et") and an unaspirated form found elsewhere (like in "s*p*ot" and "na*p*"). Thai has two phonemes, aspirated /p^h/, as in "forest" ([p^ha:]), and unaspirated /p/, as in "split" ([pa:]). Spanish has only one phoneme, unaspirated /p/, as in "but" ([ˈpeɾo]).

As a result, it's no surprise that some English speakers may have trouble clearly hearing the /p/ of Thai [pa:] or Spanish [ˈpeɾo] as "p," and not "b."

You can also understand how people from one language may have difficulty learning the sounds of a new language; a language learner must mentally form new categories. They can experience *phonemic misperception* (hearing the wrong phoneme) when this kind of listening is not yet acquired (or if it goes wrong, such as in the case of language loss after brain damage).

Eyeing the Papago-Pima language

Papago-Pima (also known as O'odham) is a Uto-Aztecan language of the American Southwest. Approximately 10,000 people speak the Papago-Pima language, mostly in Arizona. Figure 5-6 shows a brief corpus selected to show how the sounds /t/ and /ʧ/ distribute.

/t/		/ʧ/	
[ˈtaːpan]	"split"	[ˈʧihaɳ]	"hire"
[ˈtaːpaɖ]	"feet"	[ˈʧɨkid]	"vaccinate"
[ˈtaːtam]	"touch"	[ˈʧuku]	"become black"
[ˈtoha]	"become white"	[ˈkiːʧud]	"build a house for"
[ˈtoŋi]	"become hot"	[ˈʧɨgig]	"name, reputation"
[ˈwiɖut]	"swing"	[ˈʧiːwia]	"settle, establish residence"
[ˈgatwid]	"shoot"	[ˈŋumaʧ]	"liver"
[ˈtaːtaɖ]	"feet"		
[ˈwiːmt]	"help, marry"		

Figure 5-6: Selected words from the Papago-Pima language.

From the data in Figure 5-6, determine if the /t/ and /ʧ/ are separate phonemes or if they're allophones of a single underlying phoneme. If they're phonemes, show why. If they're allophones, describe their occurrence.

To do this problem, see how the sounds distribute. See if there are any minimal pairs. Look for free distribution versus complementary distribution. If the distribution is complementary, give the details of how the sounds distribute.

To solve this problem, follow these steps:

1. **Check to see if the /t/ and /ʧ/ form any minimal pairs.**

   For instance, the word [ˈtaːpan] means "split." Can you find a word [ˈʧ aːpan] anywhere that means anything? If so, you can conclude these sounds are separate phonemes (and go on your merry way); however, you'll see this is not the case. Thus, the first test of phoneme-hood fails, which means your work isn't finished.

2. **Begin to suspect allophones and check for complementary distribution.**

   You may first check along the lines of the syllable contexts, whether the sounds in question begin or end a syllable. That is, you may first be able to reason that [t] is found in one syllable position and [tʃ] in the other.

   You can quickly see that such an explanation doesn't work. For instance, a [t] is found in syllable initial position (such as in [ˈtaːpan]), as is [tʃ] (in [ˈtʃɨkid] "vaccinate"). Both [t] and [tʃ] are also found in medial position, such as in [ˈtaːtam] and [ˈkiːtʃud], and in final position, such as [ˈwiɖut] and [ˈɲumatʃ].

3. **Try other left context cues.**

   Perhaps the vowels occurring in front of the [t] and [tʃ] may provide the answer. You see that [t] can have [a] or [u] to the left of it, as in [ˈgatwid] and [ˈwiɖut], and [tʃ] can also be preceded by [i] and [a], as in [ˈkiːtʃud] and [ˈnumatʃ]. These distributions suggest some overlap.

4. **Because the left context isn't working, you can next try looking to the right of the segment.**

   Here, you find the answer. The stop consonant [t] occurs before mid and low vowels (such as /o/ and /a/), the approximant /w/, and the end of a word. However, [tʃ] is only found before the high vowels /i/, /ɨ/, or /u/.

   In other words, in Papago-Pima, [t] and [tʃ] are allophones of the phoneme, /t/. You can describe the allophones as "the palato-alveolar affricate occurs before high vowels; alveolar stops occur elsewhere."

   Congratulations! You worked out a phonological rule.

Many phonologists prefer to describe these processes more formally. Figure 5-7 shows the Papago-Pima rule.

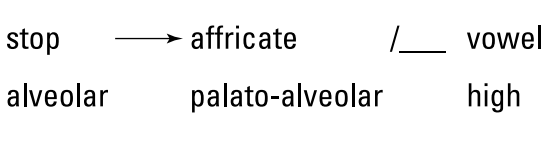


Figure 5-7: The formalized Papago-Pima rule.

Illustration by Wiley, Composition Services Graphics

Chapter 9 reviews the phonological rules for English. With these rules, you can discover how to do a narrow transcription in IPA, including which diacritics to include where. You'll be able to explain which sound processes take place in English and why, which is a highly valuable skill for language teaching and learning.

Part II

Speculating about English Speech Sounds

web extras

Visit www.dummies.com/extras/phonetics for more great Dummies content online.

In this part . . .

- Understand how consonant and vowel sounds are produced in order to classify the different sounds we use in language. Understanding sound production also helps with pronunciation.
- Differentiate between broad and narrow transcriptions, identify the purpose for each type, and begin to make your own transcriptions.
- Take a closer look at how *phonology* (sound systems and rules in languages) and *phonetics* (the study of the actual speech sounds) are related and see how together they provide a richer description of spoken language.
- Acquaint yourself with some basic phonological rules for the English language so you can make more informed transcriptions.
- Grasp the concepts of juncture, stress, rhythm, intonation, and emotion and what you need to know about them when transcribing.
- Know how to identify *prosody* (language melody) details in speech and applying what you've identified into your transcriptions.

Chapter 6

Sounding Out English Consonants

In This Chapter

- Showcasing stops
- Focusing on fricatives and affricates
- Analyzing the production of approximants
- Describing coarticulation

Producing speech is a tricky business and the exact way in which consonants are made can result in vast differences in how these sounds are heard. In this chapter, I walk you through some different types of consonant manners (stops, fricatives, affricates, and approximants), zeroing in on those mouth and throat details that make big perceptual differences in the English language.

Stopping Your Airflow

Stop consonants (sounds made by completely blocking oral airflow) are part of a larger group called *obstruents*, which are sounds formed by shaping airflow via obstruction (this group also includes fricatives and affricates). *Fricatives* are made when air is blown through a space tight enough to cause friction (or hissiness). *Affricates* are sounds that begin as a stop, then release into a fricative. Refer to Chapters 4 and 5 for more information on these types of sounds. When airflow is completely stopped, several different things can happen:

- Air can be released into the vocal tract in different ways.
- Air can flow into different regions when the sound is released.
- The duration of the closure itself can last for longer or shorter periods.

Some of these puzzling mechanics are revealed in the following sections.

Huffing and puffing: Aspiration when you need it

Aspiration is the airy event that takes place just after the burst of the articulators blasting open and before the voicing of the vowel. Aspirated voiceless stop consonants are made with an audible puff of breath. Aspiration, represented by the raised letter "h" ([^{h}]) occurs for a brief period of time starting just after the beginning of a stop consonant. To see how this works, consider what happens when you produce the word "pie."

1. **The lips close together to make the [p^{h}].**

   This is referred to as *closure.*

2. **Air pressure increases to start the [p^{h}] gesture.**

   This step refers to *oral pressure buildup.*

3. **The lips are rapidly blown apart, resulting in a typically "p"-like sound.**

   This step is also referred to as a *burst.*

4. **Because the vocal folds are open and the pressure conditions are right, a puff of air follows just after the burst.**

5. **The vocal folds start to buzz for the [aɪ] diphthong.**

If you want to feel the aspiration, place your hand just under your bottom lip while you're talking. Do you feel the air pass over your hand? That air is aspiration. Try this again and say "*p*ot." You should be able to feel the aspiration of the [p^{h}] as a puff of air hits your hand when you begin the word.

Now, try the same exercise while saying "tot" and "cot." At the beginning of these words, you also produce aspirated stops ([t^{h}] and [k^{h}]), but you may not feel much of a puff because the release is taking place farther back in your mouth. Even though you may not always feel aspiration, it's important you be able to hear and transcribe it.

Being able to work with aspiration comes with practice. In English, the voiceless stops [p], [t], and [k] are aspirated at the start of a word and at the beginning of stressed syllables. You transcribe the aspiration by adding the diacritic ([^{h}]), resulting in [p^{h}], [t^{h}], and [k^{h}]. In other contexts, [p], [t], and [k] aren't aspirated.

Table 6-1 shows you a quick overview of the rules of aspiration in English.

Table 6-1 The Rules of Aspiration in English

Context	*Examples*	*Aspiration*	*IPA*
Syllable initial	*p*ot	Strong for most speakers	[p^{h}]
	*t*ot		[t^{h}]
	*c*ot		[k^{h}]
Following an /s/	s*p*ot	None	[p]
	s*t*ock		[t]
	S*c*ott		[k]
Syllable final	ho*p*	None	[p]
	ho*t*		[t]
	ho*ck*		[k]

I use square brackets ([ ]) instead of slash marks (/ /) to mark these sounds in Table 6-1 because doing so shows narrow *phonetic* detail. The aspiration diacritic [^{h}] is included in *narrow* transcriptions of English, not broad. Aspirated stops in English occur as the result of rule–governed processes (also called *allophonic processes*).

Try saying the words in the second column and make sure you can hear the aspiration in the underlined consonants in the first row (but not in the stop consonants in the second and third row).

Declaring victory with voicing

The English voiced stop phonemes (/b/, /d/, and /g/) aren't produced with aspiration, so it may seem simple that they can be distinguished from their voiceless counterparts (/p/, /t/, and /k/). However, if you listen carefully, you should be able to tell that voicing also behaves rather differently in different environments. Take a look at Table 6-2 where you see how the amount of voicing for /b/, /d/, and /g/ changes in different environments.

Table 6-2 The Amount of Voicing in English Stops

Context	*Examples*	*Voicing*	*IPA — Narrow Transcription*
Middle of voiced phrase (VCV)	a *b*oon	Strong	[b]
	a *d*une		[d]
	a *g*oon		[g]
Sentence initial	*b*oon	Weak	[b]
	*d*une		[d]
	*g*oon		[g]
Following voiceless sound	that *b*oon	Weak	[b]
	that *d*une		[d]
	that *g*oon		[g]
Syllable final	ta*b*	Weak	[b]
	ta*d*		[d]
	ta*g*		[g]

When a voiced stop occurs between flanking voiced sounds (as shown in the first row of Table 6-2), voicing is usually strongly produced throughout the stop closure. However, in all the other cases, English [b], [d], and [g] actually aren't that strongly voiced.

The reason these weakling voiced stops (in rows 2, 3, and 4 of Table 6-2) are still heard as voiced (that is, as [b], [d], and [g]) is because other information signals listeners that a voiced sound is intended. One of these cues, voice onset time (VOT) is discussed in more detail in Chapter 14.

Another interesting way voicing is conveyed in English is by vowel length. To get an idea of how this works, concentrate on how long each word is when you say the word pairs in the following list:

tap	tab
tat	tad
tack	tag

What do you notice? You may hear that the vowel /æ/ is longer before the voiced stops /b/, /d/, and /g/ than the voiceless stops /p/, /t/, and /k/. People hear this change in *vowel* length as the voicing of the final *consonant*.

Although physical voicing may be stronger or weaker depending on the context (as shown in Table 6-2), the *feature* of voicing is abstract and perceptual. That is, the feature of voicing is in the ear of the beholder and can be signaled by various types of information.

It's possible to computer-edit versions of these words. Computer editing involves entirely removing the final consonant release and leaving only the vowel length. People still hear the (missing) final consonant voicing difference quite reliably.

Glottal stopping on a dime

If you already read Chapter 2, you discovered information about your glottis. A *glottal stop* is made whenever the vocal folds are pressed together. This process happens easily and naturally, such as whenever you cough. To make a glottal stop on command, just say "u*h*-oh" and hold the "u*h*."

Glottal stops appear in English more than people think. In London, Cockney accents are a key feature, appearing in words such as *bottle* [ˈbɒʔt̚] and, yes, *glottal* [ˈglɒʔt̚]. In North American English, glottal stops are often produced before a stop or affricate at the end of a syllable, for instance "ra*p*" or "chur*ch*."

To get a sense, try saying "ta*p*" and "ta*b*." You'll notice that the vowel in these words is longer for the second word (ending with the voiced stop, /b/), than the first, containing, /p/. Also, you may notice that you close down your glottis before you get to the final /p/ of "ta*p,*" and release no air afterwards. You probably produced [tʰæʔp].

Of course, you can pronounce "tap" in different ways. Try the varieties in Table 6-3.

Table 6-3 Different Ways to Pronounce "Ta*p*"

Pronunciation	*IPA*
With no glottal stop and no final release	[tʰæp]
With no glottal stop and final release	[tʰæpʰ]
With glottal stop and final release	[tʰæʔpʰ]
With glottal stop only	[tʰæʔ]

I give you the audio examples that are linked to each way of making the final consonant at `www.utdallas.edu/~wkatz/PFD/tap_examples.html`.

Doing the funky plosion: Nasal

In *oral plosion* (or *explosion,* when a sound is made by the articulators forced open under pressure), the articulators separate and a burst of air is released from the oral cavity. This happens for most English stops. However, when a voiced stop and a nasal occur together, as in the word "su*dden,*" something quite different happens: The air pressure built up by the stop is instead released through the nose. This process is called *nasal plosion*, which you accomplish by lowering your soft palate, also called the *velum*. Nasal plosion has the effect of producing less of a vowel-like quality for the release and more of a nasal quality. Refer to Chapter 6 for more information on oral and nasal stop consonants.

Say the word "sudden." How much of an "un" sound do you hear at the end? It shouldn't be much. Next, imagine there was an ancient poet named "*Sud Un.*" (Yes, it's a bit far-fetched, but at least it provides a different stress structure!) Say the two, side by side:

sudden Sud Un

You should be able to hear nasal plosion in *sudden*, but not in the "Un" of "Sud Un." The latter should have much more vowel quality because it's pronounced with more stress and no nasal release of the previous stop.

Notice that nasal plosion only occurs for stops that are *homorganic*, sharing the same place of articulation. This table shows the possible homorganic combinations of oral and stop consonants for English.

Oral Stops	***Nasal Stop***
/p/, /b/	/m/
/t/, /d/	/n/
/k/, /g/	/ŋ/

To put it another way, /pm/, /bm/, /tn/, /dn/, /kŋ/, and /gŋ/ are the homorganic stop/nasal combinations in English. When you say words having these combinations in English, chances are you'll use nasal plosion (as in "sudden" and "hidden"). However, when stop/nasal combinations aren't homorganic (such as /bn/ or /gn/), nasal plosion doesn't occur. You'll notice this if you say "ribbon" and "dragon," where there is no nasal release because these combinations of stop and nasal aren't homorganic.

Doing the funky plosion: Lateral

Lateral plosion involves a stop being released by lowering the sides of the tongue, instead of making an oral release by the articulator. When lateral plosion occurs, no vowel sound takes place in the syllable involved. Instead, there is more of a pure "l" sound. To get an idea, try saying these utterances, side by side, while listening to the final syllable:

Lateral Plosion	***Without Lateral Plosion***
ladle	lay dull
noodle	new dull

Depending on your accent, you may have slightly different realizations of the vowels in these expressions. There should be more of an "l" ending for the endings of the left column, and a vowel-containing ending (/əl/) for those on the right.

Tongue tapping, tongue flapping

The *tap* [ɾ] is a rapid, voiced alveolar stop used by many speakers to substitute for a /t/ or /d/. It's typically an American (and Canadian) gesture in words such as "Be*tt*y," "ci*t*y," "bu*tt*er," and "be*tt*er." (Refer to Chapter 18 where I discuss American and Canadian dialects.) I call it a *tap,* although some phoneticians refer to it as a *flap.* The difference between a tap and a flap is whether an articulator comes up and hits the articulator surface from one direction and returns (tap), or hits and continues on in the same direction in a continuous flapping motion (flap). I say we call it a tap and be done with it.

Notice that tap is shown in square brackets ([ ]) because it's an allophone in English and can't stand on its own freely to change meaning. That is, you can't say something like "Tomorrow is Fat Tuesday" [ɾəˈmɑɹo ɪz fæɾ ˌɾuzɾeɪ] where tap freely stands in for *any* /t/ or /d/.

Taps are quite important for North American English. Most Americans and Canadians replace medial /t/ and /d/ phonemes with a tap in words such as "latter" and "ladder." Forget about spelling — for spoken American English, these words often sound just the same.

Say these phrases and see how you sound:

It's the la*tt*er.

It's the la*dd*er.

If you're a native speaker of English from somewhere in North America, you may likely tap the *medial alveolar consonant* (/t/ or /d/ in the middle of a word). If you speak British English or other varieties, this isn't likely.

Having a Hissy Fit

Fricatives are formed by bringing the articulators close enough together that a small slit or passageway is formed and friction, or hissiness, results. The fricatives are copycats of many of the allophonic processes of the stops. For example, just as vowel length acts as a cue to the voicing of the following stop (as in "bi*t*" versus "bi*d*"), a similar process takes place with voiceless and voiced fricatives.

Try pronouncing the pairs in Table 6-4 and convince yourself this is the case. Notice that the /i/ in "gr*ie*ve" is longer than the /i/ in "*gr*ief" (and so forth, for the remaining pairs). This table shows examples of English word pairs having voiceless and voiced fricatives in syllable-final positions. In each case, relatively longer vowels in front cue the voiced members.

Table 6-4 Voicing Contrasts for Fricatives in Syllable-Final Position

Voiceless	*Voiced*	*IPA*	
grief	grieve	/gɹif/	/gɹiv/
teeth	teethe	/tiθ/	/tið/
race	raze	/ɹes/	/ɹez/
nation	Asian	/ˈneʃən/	/ˈeʒən/

After you've spoken the words in Table 6-4, listen carefully to the fricatives and focus on how long each fricative portion lasts. Here, you should hear another length distinction but one that's going in the opposite direction. Final voiceless fricatives are longer than final voiced fricatives. That is, the /v/ in "grie*v*e" is shorter than the /f/ in "grie*f*." See if you can hear these differences for the rest of the pairs.

This consonant duration difference is also found for stops in final position (such as "bi*t*" versus "bi*d*"). However, because stop consonants are so short, it's difficult to get a sense of this without measuring them acoustically (see Chapter 12 for more information on acoustic phonetics).

Another important point to note about the English fricatives is that four of them are *labialized*, produced with secondary action of the lips. These "lippy" upstarts include /ʃ/ and /ʒ/ (highly labialized), and /s/ and /z/ (partially labialized). For these sounds, the position of the lips helps make the closure of the fricative.

Feel the positions of your lips while pronouncing these words:

- /ʃ/: "pre*ss*ure"
- /ʒ/: "trea*su*re"
- /s/: "*s*ip"
- /z/: "*z*ip"

For these fricatives, you purse your lips to help make the sound. In contrast, for the fricatives /θ/ and /ð/ (as in "*th*ick" and "*th*is"), the placement of your lips isn't particularly important. Your tongue placed in between your teeth causes the hissiness.

The phoneme /h/ is a lost soul that needs to be given a special place of its own. Although technically classified in the IPA as a voiceless glottal fricative, its occurrence in English can be rather puzzling. (Refer to Chapter 3 for more about the IPA.) Often, it's produced without any glottal friction at all, such as "a*h*ead" or "a*h*oy there." In such cases, a weakening of the flanking vowels signal the /h/. In contrast, there may be strong frication in words such as "h*ue*" (/hju/). To add to the mix, people who produce the voiceless phoneme /ʍ/ in their dialect (as in "whip" pronounced as "hwip") are pronouncing "h" as part of an approximant (again, with no friction). So /h/ is wild and crazy, but I say you give it a home in the fricative category (as long as you remember it may not always stay put).

Going in Half and Half

Affricates are a combination of a stop followed by a fricative. English has two affricate phonemes: /ʧ/ and /ʤ/. In the IPA chart, /ʧ/ and /ʤ/ are listed as *post-alveolar* (produced by placing the tongue front just behind the alveolar ridge) because this place of articulation corresponds to the major part of the sound — namely, the fricative.

In some situations in English, a stop butts up against a *homorganic* (sharing the same place of articulation) fricative, creating situations that may seem "affricate-like." However, these instances aren't true affricates. For example, the sound /t/ can sometimes adjoin the sound /s/, as in the phrase "It seems."

However, to demonstrate that this phrase isn't a true affricate, you couldn't get away with new English expressions, such as "*ts*ello," "*ts*ow are you?" and so on, and expect anyone to think you're speaking English. This is because /ts/ can't stand alone as an English phoneme (although in other languages, such as Japanese, a /ts/ affricate phoneme is found, such as in the word "tsunami").

Shaping Your Approximants

Approximants are formed by bringing the articulators together, close enough to shape sound, but not so close that friction is created. The English voiced approximant phonemes are /w/, /ɹ/, /j/, and /l/, as illustrated in the phrase "your whirlies" /jɚ 'wɪɹliz/. In addition to this set, "hw" (written in IPA with the symbol /ʍ/) is produced by some talkers as an alternative to voiced /w/ for some words. Some pronounce "whip" or "whether" with a /w/, and others with a /ʍ/. In most forms of English, the use of /ʍ/ seems to be on the decline.

Voiced approximants partially lose their voicing when they combine with other consonants to form *consonant clusters*, lawful consonant combinations. In Table 6-5, listen as you say the approximants in the middle column, followed by the same sounds contained in consonant clusters (right column):

Table 6-5 Fully and Partially Voiced Approximants in English

Approximant (IPA)	*Fully Voiced*	*Partially Voiced*
/w/	*wh*eat	*tw*eet
/ɹ/	*r*ay	*tr*ay
/l/	*l*ay	*pl*ay
/j/	*y*ou	*c*ue

Focus on the second row in the table. Place your fingers lightly over your Adam's apple and feel the buzzing while you say the "r" in the two words. You should feel less buzzing during the "r" in "tray" than in "ray." This is because the aspiration of the voiceless stop [t^h] in "tray" prevents the approximant from remaining very voiced.

The "r" sound perhaps causes more grief to people learning English as a second language than any other. This is particularly true for speakers of Hindi, German, French, Portuguese, Japanese, Korean, and many other languages that don't include the English /ɹ/ phoneme.

Recent physiological studies show much variety in how native talkers produce this sound, although tongue shapes range between two basic patterns:

- **Bunched:** The anterior tongue body is lowered and drawn inwards, away from the front incisors, with an oral constriction made by humping the tongue body toward the palatal region. This variety is quite common in the United States and Canada.
- **Retroflex:** The tip is raised and curled toward the anterior portion of the palate.

Some clinicians use the hand cues seen in Figure 6-1 to help patients remember the bunched versus retroflex /ɹ/difference.

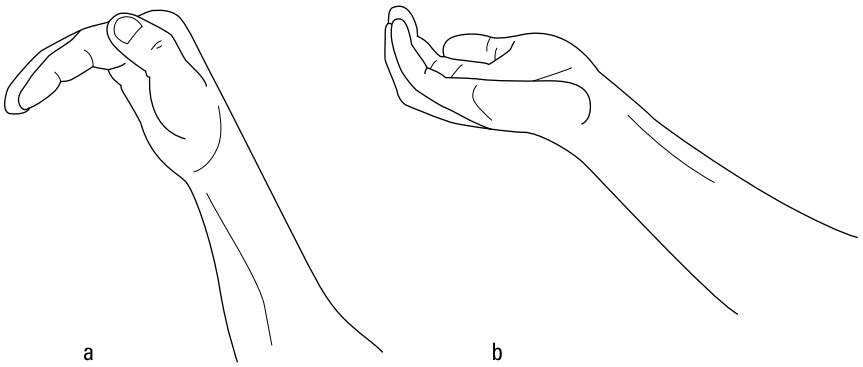


Illustration by Wiley, Composition Services Graphics

Figure 6-1: Bunched (a) versus retroflex (b) /ɹ/ hand signals.

Retroflex /ɹ/ varieties are more common in British English than in North American dialects. Also, most speakers of American and Canadian English make a secondary constriction in the pharyngeal region, as well as lip rounding behavior.

Here are some important points to know about English /ɹ/:

- /ɹ/ is a consonant.
- English also has two rhotic (r-colored) *vowels*, /ɝ/ (in stressed syllables) and /ɚ/, (in unstressed syllables); also as in "further" (/ˈfɝðɚ/).
- /ɹ/ is often called a *liquid* approximant (along with its cousin /l/) for rather odd reasons (dating back to how these sounds were used in Greek syllables).
- /ɹ/ is a relatively late-acquired sound during childhood, commonly achieved between the ages of 3 and 6 years. /ɹ/, /ɝ/, and /ɚ/ are also error-prone sounds for children, with frequent /w/ substitutions (for example, "Mister Rabbit ([ˈmɪstə ˈwæbɪt]).

Exploring Coarticulation

Speech sounds aren't produced like beads on a string. When you say a word such as "suit," you aren't individually producing /s/, then /u/, and then /t/. Doing so would sound too choppy. Instead, you produce these sounds with *gestural overlap* (overlapping movements from different key parts of your articulatory system). (Chapter 4 provides further discussion.) *Coarticulation* refers to the overlapping of neighboring sound segments. In Figure 6-2, you see an image of what that overlap looks like for the word "suit."

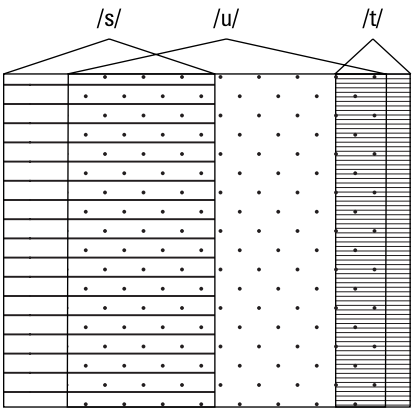


Figure 6-2: "Suit" showing sound overlap.

Illustration by Wiley, Composition Services Graphics

While the tongue, lips, and jaw are positioned to produce the frication (hissiness) for /s/, the lips have already become rounded (pursed) for the upcoming rounded vowel, /u/. This section explores some basics of coarticulation and introduces two main types of coarticulation.

Tackling some coarticulation basics

In order to better understand how coarticulation works, you need to master some important attributes. Keep in mind these general principles about coarticulation as you study more phonetics and phonology. These principles can help explain the distribution of allophones. Here are some cool things to know about coarticulation:

- All speech is coarticulated. Without it, humans would sound robot-like.
- The extent (and precision) of coarticulation differs between languages.
- Because many aspects of coarticulation are language-dependent, to some extent coarticulation must be acquired during childhood, and learned during adult second language.

However, birds (Japanese quail) have been trained to distinguish coarticulated speech sounds, suggesting that at some coarticulated processes can be accomplished on the basis of general auditory processing alone.

- Psycholinguistic research suggests children acquire coarticulation early in development.
- Coarticulation is thought to break down in certain speech and language disorders, such as apraxia of speech (AOS).

Anticipating: Anticipatory coarticulation

A "look ahead" activity is called *anticipatory* (or right-to-left) *coarticulation*. It is considered a measure of speech planning and as such is of great interest to psycholinguists (see below).

Try out anticipatory coarticulation for yourself! Say the following two phrases:

"I said suit again."

"I said seat again."

Pay special attention to when your lips begin to protrude for the /u/ in the word "s*ui*t." ***Note:*** There's no such lip protrusion for the /i/ in "s*ea*t"; this is just for comparison. Most people will begin lip rounding for the /u/ by the beginning of the /s/ of "suit," and some even earlier (for example, by the vowel /ɛ/ in the word "said"). That is, nobody waits until the /s/ is over until they begin to lip-round for the rounded vowel /u/.

These effects are important for optimizing speech speed and efficiency. The average person produces about 12 to 18 phonemes per second when speaking at a normal rate of speed. There would be no way to achieve such a rate if each phoneme's properties had to switch on and off in an individual manner (such as when using a signaling system like Morse Code). However, when speech properties are overlapped, the system can operate faster and more efficiently.

Preserving: Perseveratory coarticulation

A second type of coarticulation called *perseveratory* (or *left-to-right*) coarticulation, is also known as *carry-over*. *Perseveration* means that something continues or hangs on. In this case, it is the lingering of a previous sound on to the next. For instance, in "suit" it would be the hissiness of the /s/ carrying over to the beginning of the vowel /u/, or the rounding of the /u/ continuing on and influencing the final /t/. Perseverative coarticulation is a measure of the mechanical/elastic properties of the speech articulators, instead of planning.

One way to remember perseverative coarticulation is to think about the role of this property in complex speech, such as *tongue twisters*. A property common to tongue twisters throughout the world is that they have phonemes with similar features in close proximity. The hope is that you'll have carryover effects from a sound you just made as you attempt to produce an upcoming sound with similar properties. Actually, if you begin thinking about it too much, you might then develop anticipatory problems as you desperately thrash around trying to keep the proper phonemes in mind. This is evident in the saying "She sold seashells by the seashore."

You may end up saying "She *sh*old" as you carry over from the initial /ʃ/ of "she" to the target /s/ of "*s*old."

How far can you stretch it? Look-ahead coarticulation in English and French

English speakers only show anticipatory lip rounding over the course of a syllable or so. For example, in the word "suit," there's reliable anticipation of lip rounding for /u/ during the initial /s/. However, if someone begins to load up the front of the syllable such that more consonants intervene between the /s/ and the /u/, anticipatory lip-rounding gradually diminishes (such as more in "*s*uit" than in "*sp*ool").

What about other languages where lip rounding plays a more important (phonemic) role?

A now classic study asked six French speakers to produce a series of tongue-twister type expressions to see how far lip protrusion may extend. The subjects had a photocell attached to their upper lips as they said things like "une sinistre structure (a sinister structure)." The researchers measured when (and how far and how fast) the upper lip began to protrude for the rounded vowel /y/ in the upcoming word /strykˈtyr/.

The researchers found that notable lip protrusion for the /y/ of /strykˈtyr/ began as early as four to six consonants before.

In a follow-up experiment, listeners were given gated segments of the consonant clusters prepared by a waveform-editing program and asked to detect whether the segment was taken from an utterance that was before a rounded vowel. Listeners could do this at better than 50 percent accuracy by up to four consonants before the rounded vowel, suggesting that long-ranging labial coarticulation can be accurately tracked by listeners.

These experiments provided early evidence that when it comes to coarticulation, one size does not fit all: Languages which emphasize certain sound features (such as lip-rounding) in their sound inventories (such as French) have different coarticulatory features for these sounds than do languages (such as English) that don't.

Chapter 7

Sounding Out English Vowels

In This Chapter

- Searching for (IPA) meaning in all the right places
- Hearing vowels in full and reduced forms
- Switching between British and North American English vowels
- Keeping track of vowel quality over time

Vowels are a favorite subject of phoneticians because they play such an important role in perception, yet they pose so many mysteries about how speech is perceived and produced. Some vowels are quite easy to transcribe; some remain difficult. In this chapter, I highlight the commonalities among English vowels by describing the group's tense and lax characteristics. I also talk about rhoticization (also referred to as r-coloring), which is important for many applications, including the description of various English accents and understanding children's language development.

Cruising through the Vowel Quadrilateral

Making vowels is all about the tongue, lips, and jaw. However, the final product is *acoustic* (sound related), not *articulatory* (mouth related). Phonetics texts typically start out with articulatory instructions to get people started, but it becomes important to transfer this information to the ear — to the world of auditory information.

In articulatory phonetics, vowels are studied using the *vowel quadrilateral*, a trapezoid-like diagram that classifies vowels according to tongue height, advancement (front-back positioning), and lip rounding.

This section focuses on moving your tongue to known target regions and consciously getting used to what these regions sound like. In this way, sound anchors become familiar landmarks as you cruise through the land of vowels.

Looking up vowel sounds

Are you a book person, an Internet person, or both? The tools available for looking up sounds are becoming more extensive and convenient. As your transcription skills improve, you may find yourself wishing to compare pronunciations of certain words now and then. At such times, you may notice quite a bit of difference among various reference sources because vowels can be transcribed in a number of ways, depending on the exact needs of the transcription.

In the following, I compare some dictionary and Internet sources for the English words "cheap" and "chip" (American pronunciation) to see how they're transcribed. This table can give you an idea of how transcription is handled in a range of available sources.

Source	"cheap"	"chip"
English Pronouncing Dictionary (EPD)	/ʧi:p/	/ʧɪp/
Longman Pronunciation Dictionary (LPD)	/ʧi:p/	/ʧɪp/
Oxford Dictionary of Pronunciation for Current English (2003)	/ʧi:p/	/ʧɪp/
American Heritage Dictionary 5th Edition (2011)	(chēp)	(chĭp)

I selected these sources because they're some of the most authoritative print dictionary sources. The first three sources provide both British and American English pronunciations. Unlike the conventions used by most phonetics teachers (and used in this book), these first three sources are also more detailed in that they show both length and quality features for the "cheap/chip" distinction. That is, there are two related phonetic factors that contribute to the vowel difference:

- **Quality:** The formants are different in the two words, requiring different symbols /i/ versus /ɪ/
- **Length:** The vowel is longer in *cheap* than "chip," requiring a length diacritic for *cheap* /i:/

The Oxford Dictionary uses an even larger set of symbols and gives more details than the first two sources.

A different approach has been taken in the American Heritage, where an alternative to the International Phonetic Alphabet (IPA) was used, presumably to be more user friendly. This, however, isn't as handy or reliable for some people who've already learned the IPA.

This book follows a convention used by many phonetics instructors and transcribes vowel quality, assuming that length can be implied. Therefore, I transcript "cheap" as /ʧip/ and "chip" as /ʧɪp/. As long as you remember that quality and length go together, you should be able to appreciate other transcriptions (such as EPD and LPD) when you use them.

Sounding out front and back

Sound-based descriptions are especially important for vowels. For this reason, phoneticians have long relied on perceptual descriptions of vowels.

For instance, front vowels were frequently called *acute* because they're perceptually sharp and high in intensity. These vowels also trigger certain sound changes in language (notably, palatalization) and involve active tongue blade (coronal) participation. In contrast, back vowels were called *grave* because they have dull, low intensity.

You have made front vowels, but you have probably not spent that much time attending to what the vowels *sound* like. So here, you tune in to the sounds themselves. Begin by making an /i/ as in "heed," then move to the following, one by one:

- /ɪ/ as in "h*i*d"
- /e/ as in "h*aye*d"
- /ɛ/ as in "h*ea*d"
- /æ/ as in "h*a*t"

You hold your tongue in a certain position for each vowel (although there is some wiggle room), and the tongue position need not be exact. Also, each vowel position can blend somewhat into the position of the next.

Now, try saying them all together in a sequence, /i ɪ e ɛ æ/. Notice that the vowels are actually in a continuum. Unlike consonants, vowels are made with the tongue relatively free in the articulatory space and the shaping of the whole vocal tract is what determines the acoustic quality of each sound.

Now try the same listening exercise with the back vowel series, beginning with /u/ as in *who'd*, and proceeding to the following:

- /ʊ/ of "h*oo*d"
- /o/ of "h*o*pe"
- /ɔ/ of "l*aw*"
- /ɑ/ of "d*o*g"

In the back vowel series, you pass through the often-confused /ɔ/ and /ɑ/. There are many dialectal differences in the use of these two vowels. For instance, in Southern California (and most Western United States dialects), most talkers pronounce "cot" and "caught" with /ɑ/. In Northern regions, say Toronto, talkers use /ɒ/ for both words. This vowel /ɒ/ is a low back vowel similar to /ɑ/ but produced with slight lip rounding. However, elsewhere in the States (especially in the Mid-Atlantic States) talkers typically produce "cot" with /ɑ/ and "caught" with /ɔ/. You can easily tell the two apart by looking at your lips in a mirror. During /ɑ/, your lips are more spread than in /ɔ/, and in /ɔ/ the lips are slightly puckered. Compare your productions with the drawings in Figure 7-1.

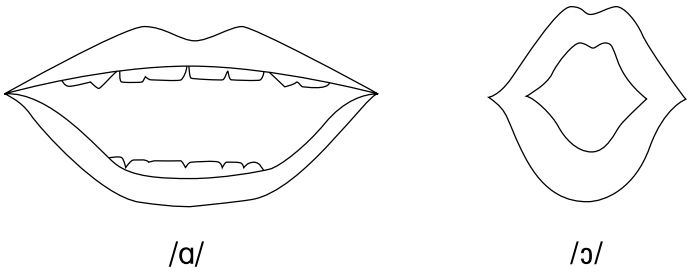


Figure 7-1: Lip positions for /ɑ/ versus /ɔ/.

Illustration by Wiley, Composition Services Graphics

The point here is to be able to *hear* such differences. Try moving from an /o/ as in "hope" to an /ɔ/ in "l*aw*" to an /ɑ/ in "f*a*ther." Now try this while adding lip rounding to the /ɑ/. You should hear its quality change, sounding more like /ɔ/. As you get more fine-grained in your transcriptions, you need to be able to distinguish vowels better, including whether lip rounding occurs as a secondary articulation.

Stressing out when needed

In English, *stress* refers to a sound being longer, louder, and higher. Stress is a *suprasegmental* property, meaning it affects speech units larger than an individual vowel or consonant. I also discuss stress in Chapters 10 and 11.

In English, the amount of stress a syllable receives influences vowel quality. Stressed syllables tend to have a full vowel realization, while unstressed syllables have a centralized, reduced quality. Sometimes there is a more complicated situation, where a full vowel will appear in fully stressed syllables, but whether a vowel is reduced in unstressed syllables depends on the particular word involved. Take a look at Table 7-1, and try saying the English words.

Table 7-1 Vowels in Different Stress Conditions

Vowel	*Fully Stressed*	*Unstressed, Not Reduced*	*Unstressed, Possibly Reduced?*
/i/	ag*ree*	r*e*active	*re*sourceful
/u/	ref*ute*	ref*u*tation	Herc*u*les
/aɪ/	re*cite*	*ci*tation	*recit*ation

You should have nice, full /i/, /u/, and /aɪ/ vowels in the words of the second column of Table 7-1. This should also be the case for the words in the third column.

The most variable response is the fourth column. The vowels produced here depend on your accent. These words contain unstressed syllables that some speakers produce with a fully realized vowel quality (for example, /ɹi/ for the first syllable of "*re*sourceful") while others use a reduced vowel instead (such as /ɹə/). If your vowel is a bit higher toward /ɪ/, it may qualify for being /ɨ/ (called barred-I in IPA), as is frequently heard in American productions of words such as "dish*es*" and "rich*es*."

By the way: If you find yourself almost forgetting what you normally sound like, please remember these rules:

- **Use a carrier phrase.** A *carrier phrase* is a series of words you place your test word into so that it's pronounced more naturally. For example, "I said ___ again."
- **Have one or two repetitions and then move on.** Natural speech is usually automatic and not consciously fixated on. If a word or phrase is repeated over and over, this natural, automatic quality may be lost.

Coloring with an "r"

Whether or not people produce an "r" quality in words like "*fur*ther," "fath*er*," and "s*ir*" is a huge clue to their English accent. Most speakers of North American English produce these vowels with *rhoticization.* This term, also referred to as *r-coloring,* means that the vowel (not the consonant) has an "r"-like sound. If the vowel is stressed, as in "*fur*ther" or "*sir*," then you use the mid-central stressed vowel /ɝ/ symbol for transcription. For unstressed syllables, such as the "er" of "fath*er*," you instead use the IPA symbol schwar /ɚ/.

R-coloring is a perceptual quality that can be reached in a number of ways. R-coloring demonstrates the property of *compensatory articulation*, that a given acoustic goal can be reached by a number of different mouth positions.

R-coloring can differ substantially among individual speakers. Some make a *retroflex* gesture, putting the tip of the tongue against the rear of the alveolar ridge, while others hump the tongue in the middle of the mouth, sometimes called American *bunched r.* These vowel gestures are very similar to the consonant /ɹ/ in English and are described in detail in Chapter 6.

A useful series of r-colored vowels can be elicited in the context /fVɹ/ where V stands for a vowel. Table 7-2 contains many of these items and some others, including common North American English and British English words. Try these words out and see how much rhoticization (r-coloring) you use.

Table 7-2 Vowel R-Coloring in Three Varieties of English

IPA	*Example*	*American English*	*Canadian (similar to AE unless indicated)*	*British English (RP)*
/i/	seer	/siɹ/ or /siə˞/	…	/siə/
/ɪ/	fear	/fɪɹ/	…	/fɪə/
/e/	payer	/peɹ/ or /peə˞/	…	/peə/
/ɛ/	fair	/fɛɹ/	/fer/	/fɛə/
/ɜ˞/	fur	/fɜ˞/	…	/fɜː/
/ʊ/	poor	/pʊɹ/	…	/pʊə/
/ɔ/	sore	/sɔɹ/	…	/sɔə/
/ɑ/	far	/fɑɹ/	…	/fɑː/
/aɪ/	fire	/faɪɹ/	/fʌɪɹ/	/faɪə/
/aʊ/	flower	/flaʊɹ/	…	/flaʊə/
/ɔɪ/	foyer	/fɔɪə˞/	…	/fɔɪə/

Different transcription systems may be used for non-rhotic forms of English, such as commonly found in parts of the United Kingdom, Ireland, South Africa, and the Caribbean. I give more detail on different accents in Chapter 18.

A symbol used to describe the central nonrhotic (stressed) vowel is /ɜ/ (reversed epsilon). You can find this vowel in Received Pronunciation (RP) British for words such as “fur” and “bird.”

Neutralizing in the right places

The vowels /o/ and /i/ make predictable changes in particular environments. Phoneticians have adopted conventions for transcribing these patterns. For example, take a look at these transcriptions (GAE accent):

sore	/sɔɹ/
selling	/ˈsɛlɪŋ/

Beginning transcribers are often puzzled as to why /ɔ/ is used in "s*o*re" (instead of /o/), and why /ɪ/ is used before /ŋ/ in words that end with -ing, such as "selling." The answer is that vowels are affected by their surrounding consonants. These effects are more pronounced with certain consonants, especially the liquids (/ɹ/ and /l/) and nasals (/m/, /n/, and /ŋ/). This results in *neutralization*, the merger of a contrast that otherwise exists. For example, /o/ and /ɔ/ sound quite distinct in the words "b*oa*t" and "b*ough*t" (at least in GAE). However, before /ɹ/ these vowels often neutralize, as the /ɹ/ has the effect of lowering and fronting the /ɔ/ toward the /o/. Front vowel examples include "t*ie*r" and "p*ie*r" (pronounced with /ɪ/). The same process can take place before /l/. Examples include "pill" and "peel," both produced as /pɪɬ/ in some accents.

Say "runn*ing*." Do you *really* make a tense /i/ as in "b*ea*t" during the final syllable? Probably not. For that matter, you're probably not making a very pure /ɪ/ either. You're neutralizing, making something in-between. To label this sound, phoneticians lean toward the lax member and label it /ɪ/. Thus, /ˈɹʌnɪŋ/.

It's the same principle with "s*o*re." You're probably not using a tense /o/, such as in "b*oa*t." Listen closely! The closest vowel that qualifies is /ɔ/, even though its quality is different when rhotic.

Tensing up, laxing out

This tense versus lax vowel difference is important for a number of applications in language instruction and clinical linguistics. Specifically, the tense-lax difference indicates whether a vowel can stand alone at the end of a stressed syllable (*tense*), or whether the syllable must be closed off by a consonant at the end (*lax*). Many languages (such as Spanish) don't have any of the English lax vowels, and native speakers will therefore have difficulty learning them when studying English as a second language.

Take a look at these word pairs and pronounce them, one by one.

"beat" versus "bit"

"bait" versus "bet"

"Luke" versus "look"

Can you hear a systematic change in the sound of each pair? The first member of each pair is tense, and the second member, lax. This distinction was originally thought to result from how the vowels were made, muscularly. However, these differences are now understood as relating to English *phonology* (system of sound rules). Refer to Table 7-3 for examples.

Table 7-3 Distribution of English Tense and Lax Vowels

	Vowel	*Stressed Open Syllable*	*Closed Syllable*
Tense	/i/	bee /bi/	beat /bit/
Lax	/ɪ/	bih /bɪ/ (not a real word)	bit /bɪt/

The tense vowel /i/ can appear in a stressed open syllable word such as "b*ee*," or in a syllable closed with a consonant at the end, such as "b*ea*t." If you try to leave a lax vowel in a stressed open syllable (such as the made-up word "bih"), you end up with something very un-English-like. You can pronounce such a word, but it will sound like something from another language. The same is true with /ɛ/, /æ/, /ʊ/, and /ʌ/. You can't really go around saying "That is *veh*. I appreciate your *geh* very much."

Because of this restriction of not being able to appear in stressed open syllables, /ɪ/, /ɛ/, /æ/, /ʊ/ and /ʌ/, as in "h*i*d," "h*ea*d," "h*a*d," "h*oo*d," and "m*u*d" are called the *lax* vowels of English. Most phoneticians consider the vowels /ɑ/, /i/, /u/, /e/, and /o/ to be the *tense* vowels. These vowels are produced more at the edges of the vowel space (less centralized) than their lax counterparts. You can hear the difference between these tense vowels and their corresponding lax member in the pairs /i/ and /ɪ/, /e/ and /ɛ/, and /u/ and /ʊ/. If you say these in pairs, you should be able to hear both a difference in quality and quantity (with the lax member being shorter in duration). The /ɑ/ and /o/ tense vowels don't really have a lax member to pair up with (oh, well — somebody has to stay single!).

RP: Received from *whom*, exactly?

RP, the abbreviation for the *Received Pronunciation*, is a prestigious accent spoken by approximately 2 to 3 percent of people in the United Kingdom. This includes the royal family and some members of the government and the media. The term *RP* is usually credited to phonetician Daniel Jones, although the usage can be traced back earlier. The idea of *received* means approved, like *received wisdom.* RP is often associated with the south of England, but is actually a blend of speech from East Midlands, Middlesex, and Essex. Historically, RP was common at Oxford University. As more British families sent their children to the British public schools during the 19th century (similar to American private schools), this accent took hold as a symbol of access to this world of education and power. Currently, RP is admired in some circles and viewed negatively in others.

It's important to remember that a person can speak grammatically correct English with a Standard English dialect (often called SE, for Britain) without speaking the RP accent. Because the RP accent is well documented and is used in dictionaries, it's frequently referenced here and in other phonetics texts. Chapter 18 gives more on this dialect (and other British varieties).

Most forms of British English have one more lax vowel than American English, /ɒ/ called *turned script a* in IPA. This is an open, back rounded vowel, as in RP "cod" and "common." It can't appear in stressed open syllables and is lax.

Sorting the Yanks from the Brits

Phoneticians focus on the sound-based aspect of language and don't fret about the spelling, syntax (grammar), or vocabulary differences between North American and British varieties. This helps narrow down the issues to the world of phonetics and phonology.

In terms of vowels, you need to consider other issues than just the presence or absence of rhotics. There are quality differences in monophthongs as well as different patterns of diphthongization depending on which side of the pond you live. These sections take a closer look at these differences.

Differentiating vowel sounds

For front vowels (ranging from /i/ to /æ/), both North American English and British English have sounds spaced in fairly equal steps (perceptually). You should be able to hear this spacing as you pronounce the words "heed," "hid," "head," and "had." Try it and see if you agree.

Things get testy, however, with the vowel /e/. English /e/ is transcribed as /eɪ/ by many phoneticians (especially in open syllables) because this vowel is typically realized as a diphthong, beginning with /e/ and ending higher, usually around /ɪ/. This is shown in a traditional vowel quadrilateral (Figure 7-2a). Overall, the amount of diphthongal change for American /eɪ/ is less than that found for the major English diphthongs /aɪ/, /aʊ/, and /ɔɪ/.

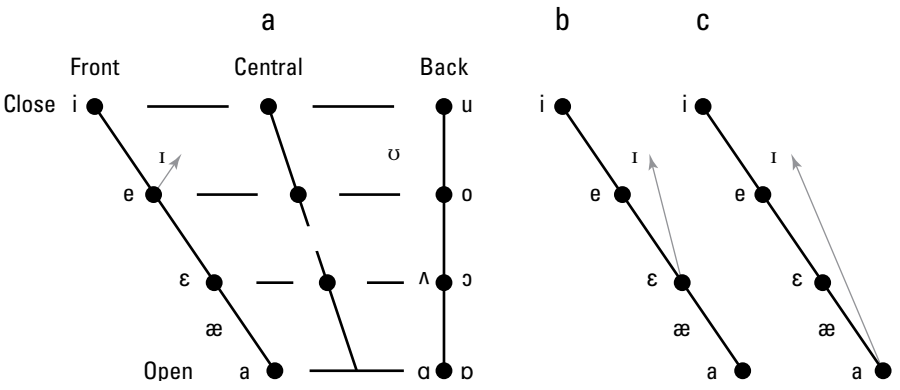


Illustration by Wiley, Composition Services Graphics

Figure 7-2: Vowel quadrilateral showing different offglides used for varieties of GAE (both a and b) and British English (c).

However, talkers vary with respect to where they really start from. Fine-grained studies of American English talkers suggest that many people start from lower vowel positions, producing words like "great" as /gɹɛɪt/. The trajectory of this diphthong is shown in Figure 7-2b. Forms of English spoken in the United Kingdom have different trajectory patterns. The direction of the /e/ diphthong for RP is similar to the direction of the GAE /eɪ/, but extends slightly further (not shown in figure).

Other British dialects have larger diphthong changes, including London accents sometimes called *Estuary English* (see Chapter 18). These upstarts (named for people living around the Thames, not birds), produce /seɪ/ "say" sounding more like /saɪ/. A panel showing the diphthong trajectories of this accent is shown in Figure 7-2c. Not to be outdone, the Scots arrive at a vowel like the Japanese, doing away with a diphthong altogether and instead producing a high monophthong that can be transcribed [e]:

"Which way (should we go to Lochwinnoch?) [ʍɪtʃ we:] . . .

There are even more North American versus British differences in the mid and back vowels. Starting with the mid vowel /ʌ/, British speakers produce the vowel lower than their North American counterparts. This is likely due to the fact that British talkers distinguish words like "bud" and "bird" by distinguishing between low /ʌ/ and the higher mid-central vowel /ɜ/. However, North American talkers use a rhotic distinction (/ʌ/ versus /ɝ/) and don't require this height separation.

North American talkers show regional differences among the back vowels, particularly for the notorious pair /ɑ/ and /ɔ/. The tendencies are either to merge the two toward /ɑ/ (Southern California) or closer to /ɔ/ (Northern American dialects). Most speakers of British English have added another vowel to the mix: high back rounded /ɒ/.

Table 7-4 shows some examples of these British back vowel distinctions so you can get grounded in the differences. This may be especially helpful if you're interested in working on accents for acting, singing, or other performance purposes. (I also include URLs where you can listen to audio files.)

Table 7-4 English Back Vowels – British SE

Back Vowel	*Examples*	*IPA*	*URL to Listen to*
/ɑ/	balm	/bɑm/	www.utdallas.edu/~wkatz/PFD/balm.wav
/ɒ/	bomb	/bɒm/	www.utdallas.edu/~wkatz/PFD/bomb.wav
/ɔ/	bought	/bɔt/	www.utdallas.edu/~wkatz/PFD/bought.wav

These differences provide an insight into the challenges facing people trying to master new accents. Namely, it's difficult moving from an accent with fewer distinctions (such as no difference between /ɑ/ and /ɔ/) to an accent with more distinctions. This is not only because the learner must use more sounds but also because the distribution of these sounds isn't always straightforward.

For example, British RP accent uses an /ɑː/ sound for many words that American English uses an /æ/ for. For instance, "glass" and "laugh" (/glɑːs/ and /lɑːf/). However, speakers of RP pronounce "gas" and "lamp" the same as in GAE, with /æ/. Thus, a common mistake for GAE speakers attempting RP is to overdo it, producing "gas" as /gɑːs/. Actually, there is no easy way to know which RP words take /ɑː/ and which take /æ/, except to memorize.

Notice that it's not as tricky to go in the opposite direction, from more accent distinctions to less. For example, a British RP speaker trying to imitate a California surfer could simply insert an /ɑ/ vowel for "b*om*b," "b*al*m," and "b*ou*ght" and probably get away with it. But could that British person actually surf?

English has a diphthongal quality to the tense vowels /e/, /i/, /o/, and /u/, particularly in open syllables. For this reason, these vowels are often transcribed /eɪ/, /ij/, /oʊ/, and /uw/ (see also Chapter 2).

Dropping your "r"s and finding them again

Rhotic and non-rhotic accents are a bit more complicated than is indicated in the "Coloring with an 'r'" section, earlier in the chapter. Many of the non-rhotic accents (they don't pronounce an "r" at the end of a syllable) express an /ɹ/ under certain interesting circumstances.

A *linking-r* occurs if another morpheme beginning with a vowel sound closely follows nonrhotic sounds. This is typical of some British accents, but not American Southern States. Here are a couple examples.

Example Word	***British SE***	***American Southern States***
care	/keə/	/keə/
care about	/ˈkeəˠ əbaʊt/	/ˈkeəʔ əbaʊt/

A similar-sounding process is *intrusive-r*, the result of sound rules trying to fix things that really aren't broken. For these cases, such as *law-r-and-order,* an "r" is inserted either to fix the emptiness *(hiatus)* between two vowels in a row, or to serve as a *linking*-r that was never really there in the first place (for example, if "tuna oil" is pronounced "tun*er* oil"). Table 7-5 shows some examples. I also include URLs where you can listen to audio files.

Table 7-5 Examples of Linking-r

Phrase	*IPA*	*URL*
Australia or New Zealand	/ɒsˈtɚ˞ɪlɪə˞ ɔ:nju: ˈzi:ln̩d/	www.utdallas.edu/~wkatz/PFD/Linking_R1.wav
There's a comma after that.	/ðəzə ˈkɒmə˞ ɑ:ftə θæt/	www.utdallas.edu/~wkatz/PFD/Linking_R2.wav
Draw all the flowers	/drɔ:ɹ ˈɔ:l ðə flaʊəz/	www.utdallas.edu/~wkatz/PFD/Linking_R3.wav

Noticing offglides and onglides

There are a number of different ways to describe the dynamic movement of sound within a vowel. One way, as I describe in Chapter 2, is to classify vowels as monophthongs, diphthongs, or triphthongs. This description takes into account the number of varying sound qualities within a vowel. Phoneticians also note which part of the diphthongs (the end or the beginning) is the most prominent (or unchanging). This distinction is commonly referred to as offglides and onglides:

- **Offglides:** If the more prominent portion is the first vowel (as in /aɪ/), the second (nonsyllabic) part is the *offglide*. This idea of an offglide also provides a handy way to mark many types of diphthongs that you may find across different accents. For instance, in American Southern States accents, lax /æ/ becomes /ɛə/ or /eə/. That is, they are transcribed including a /ə/ offglide. Table 7-6 shows some examples with URLs to audio files.

Table 7-6 Vowels Produced with an Offglide

Example Word	*IPA*	*URL*
lamp	/leəmp/	www.utdallas.edu/~wkatz/PFD/lamp-offglide.wav
gas	/geəs/	www.utdallas.edu/~wkatz/PFD/gas-offglide.wav

Some phoneticians denote an offglide with a full-sized character (such as /eə/), while others place the offglide symbol in *superscript* (such as /eᵊ/).

- **Onglides:** An *onglide* is a transitional sound in which the prominent portion is at the end of the syllable. These sounds begin with a constriction and end with a more open, vowel quality.

  An example of an onglide in English would be the /j/ portion of /ju/. Some phoneticians treat this unit as a diphthong, while a more traditional approach is to consider this syllable a combination of an approximant consonant followed by a vowel.

Doubling Down on Diphthongs

American English and British English accents have in common a set of three major diphthongs, /aɪ/, /aʊ/, and /ɔɪ/. These are called *closing diphthongs*, because their second element is higher than the first (the mouth becomes more closed). You can see the three major diphthongs (similar in GAE and British English) in Figure 7-3a, and a minor diphthong (found in British English) in Figure 7-3b. The /aɪ/, /aʊ/, and /ɔɪ/ diphthongs are also called *wide* (instead of narrow) because they involve a large movement between their initial and final elements.

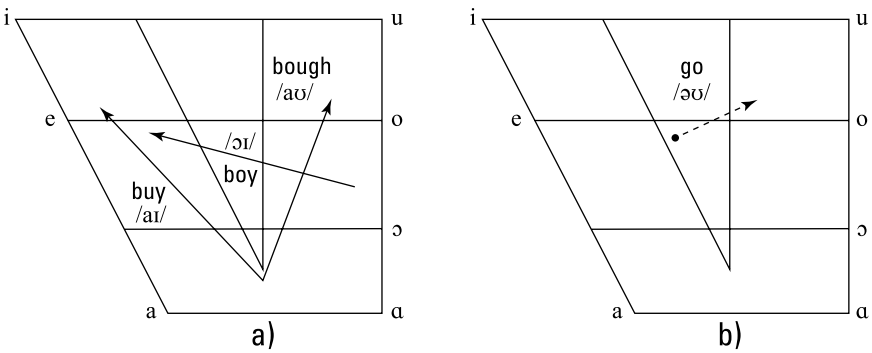


Illustration by Wiley, Composition Services Graphics

Figure 7-3: Diphthongs found in both GAE and British English (a), and in only British English (b).

Considering first /aʊ/, as in "cow," a similar trajectory is seen in BBC broadcaster English as in GAE. The /aʊ/ diphthong is also called a *backing* diphthong because posterior tongue movement is involved when moving from /a/

to /ʊ/. As may be expected, there are many variants on this sound, especially in some of the London accents (which can sound like gliding through /ɛ/, /ʌ/, /u/ or /æ/, /ə/, and /ʊ/).

The /aɪ/ sound is a *fronting* diphthong. An important thing to remember about this sound is that few talkers will reach all the way up to a tense /i/ for the offglide; it's usually /ɪ/. A second fronting diphthong found in British English and American English accents begins in the mid back regions. This is the diphthong /ɔɪ/, as in "b*oy*," "Fl*oy*d," and "*oi*l".

An interesting diphthong found in British accents (but *not* in GAE) is the closing diphthong /əʊ/. Look at the dotted line in Figure 7-3b. This sound is found in place of the GAE tense vowel /o/. Because it doesn't have much of a sound change, it would qualify as a narrow diphthong. Table 7-7 shows some examples. You can also check out the audio files.

Table 7-7 How to Say "o" in GAE and RP

Example Word	*GAE*	*RP*	*URL*
go	/goʊ/	/gəʊ/	www.utdallas.edu/~wkatz/PFD/go-GAE-RP.wav
so	/soʊ/	/səʊ/	www.utdallas.edu/~wkatz/PFD/sew-GAE-RP.wav

Lengthening and Shortening: The Rules

This section concentrates on *vowel length*, namely how a given vowel's length changes as a function of context. Such context-conditioned change is called *allophonic variation* (see Chapter 5 for more information).

If you're an English speaker, you naturally carry out at least three subtle timing changes for vowels when you speak. Here, I note these processes formally as *rules*. This information can come in handy if you teach English as a second language, compare English to other languages, or engage in any work where you need to be able to explain what the English sound system is doing (instead of, say, stamping your feet and saying "because that's just how it is!").

Vowel Inherent Spectral Change (VISC): See you in court?

Monophthongs don't have as much quality change as diphthongs and triphthongs. However, exciting new research by Professors Terrance Nearey, Peter Assmann, and others has shown that in many languages (particularly those with large vowel inventories) monophthongs also demonstrate substantially changing sound patterns. This information is called *vowel inherent spectral change (VISC).* It seems to be important for human vowel perception, affecting speech development, second language learning, and dialect change.

Research by Professor Geoffrey Stewart Morrison has shown that VISC can provide useful information for forensic voice comparison. Although this work is still in an early stage, the goal would be to boost speaker identification using VISC from recordings of a subject's speech.

Check out each rule and its examples:

- **Rule No. 1:** Vowels are longest in open syllables, shorter in syllables closed by a voiced consonant and shortest when in syllables closed by a voiceless consonant. For example:

  "bay" (/beɪ/)

  "bayed" (/bed/)

  "bait" (/bet/)

- **Rule No. 2:** Vowels are longer in stressed syllables. For example:

  "re*peat*" (/ɹəˈpit/

  "to *re*peat" (/ˈɹipit/)

  Here, "peat" (/pit/) should sound longer in the first than the second example.

- **Rule No. 3:** Vowels get shorter as syllables are added to a word (up to three syllable-words). For example:

  "zip" (/zɪp/)

  "zipper" (/ˈzɪpɚ/)

  "zippering" (/ˈzɪpɚɪŋ/)

Chapter 8

Getting Narrow with Phonology

In This Chapter

- Digging into phonology
- Sorting out types of transcription
- Getting a sense of rule ordering and morphophonology

As you study phonetics, many of the IPA symbols and the sounds of English will become warmer and cozier, as you become more familiar with them. You can look at symbols, such as /æ/ and /ʃ/ and know they represent sounds in the words "c*a*t" and "*sh*out." To help you be more comfortable, you need a firm grasp of the relationship between phonetics and phonology, which allows you to move between broad and narrow transcription. This chapter helps clarify how phonetics and phonology are related, which can help you take your transcriptions to the next level.

Phonetics is the study of the sounds of language. Phonetics describes how speech sounds are produced, represented as sound waves, heard, and interpreted. Phonetics works hand-in-hand with *phonology*, the study of the sound systems and rules in language.

Phonologists typically describe the sound processes of language in terms of *phonological rules*, patterns that are *implicit* (naturally understood) by speakers of the language. For example, a speaker of English naturally (implicitly) nasalizes a vowel before a nasal consonant, as in "run" ([ɹʌ̃n]) and "dam" ([dæ̃m]). English speakers nasalize a vowel even for a nonsense words, such as "zint" ([zĩnt]) and "lemp" ([lẽmp]).

If you're a native English speaker, you'll also nasalize the vowels in these examples. Go ahead and speak these nonsense words ("zint" and "lemp"). Word meaning has nothing to do with it; it's a sound thing!

Part of knowing a language entails you understand and use its phonology, processes that can be described so you'll be able to incorporate information about language sound rules into your transcriptions. The following sections explain the main kinds of transcriptions and how they differ.

Distinguishing Types of Transcription

Phonetic transcription uses symbols to represent speech sounds. However, depending on your need, you can transcribe in many different ways. A transcription can look quite different based on whether you'll use it for theoretical linguistics, language teaching, speech technology, drama, or speech and language pathology. Here are some important distinctions used to classify the main types of transcriptions.

Impressionistic versus systematic

The transcriber's knowledge can play a key role in two main types of transcription classifications. They are:

- **Impressionistic:** An *impressionistic* transcription occurs when you, as the transcriber, have minimal knowledge of the language, dialect, or talker being worked with. As such, you'll use your minimal experience to make judgments about the incoming sounds. An example would be somebody trying a first transcription of a complex African language. In such a situation, the transcriber could only hope to describe the new language in terms of the categories of his or her native language. The results probably wouldn't be very accurate because the transcriber wouldn't know which details would turn out to be important.
- **Systematic:** In contrast, if you, as the transcriber, are well trained in phonetics and had made several passes over the new language, you can note important details. This transcription would be *systematic*, reflecting the structure of the language under description.

Impressionistic and systematic are therefore endpoints on a continuum. The more detailed and accurate the transcription, the more it moves from impressionistic to systematic.

Broad versus narrow

Transcription can also be classified as simple or detailed, as the following explains:

- **Broad**: The simpler your transcription (with the less phonetic detail), the more broad it is. *Broad* transcription has the advantage of keeping the material less complicated. Although a broad transcription is sufficient for many applications and you can complete a broad transcription with less phonetic training, you basically get what you pay for. If you want to later go back to these transcriptions and reproduce the fine details, you'll probably be out of luck.

- **Narrow:** A maximally *narrow* transcription indicates all the phonetic detail that is available and relevant. Completing a narrow transcription requires more training than simply knowing IPA characters: You must know something about the phonology of the language and the diacritics typically used to designate *allophones* (contextually-related sound variants). Narrow transcriptions offer substantial detail, useful for scientific and technical work. Making sure that such transcriptions don't become needlessly cluttered is important; otherwise, readers may have a nightmare getting through it.

Like the preceding section (on impressionistic and systemic dichotomy), the broad and narrow contrast can be best thought of in terms of a continuum. That is, a transcription can range from broad to narrow.

Capturing Universal Processes

Just as phonetics has a universal slant (to describe the speech sounds of *language* — as in all of the languages of the world), phonology also seeks to describe the sound processes of all the world's languages. This emphasis on universal goals has affected how phonetics and phonology are taught worldwide. For example, whereas phonetics and phonology used to be taught predominantly within the auspices of particular language and literature departments (such as English and the Slavic languages), they're now frequently integrated with linguistic, cognitive, and brain sciences because of the assumption that speech and language are universal human properties.

Getting More Alike: Assimilation

One of the most universal of sound phonological rules in language is *assimilation*, when neighboring sound segments become more similar in their production. They're frequently called *harmony* processes.

At a physiological level, you can describe assimilation as *coarticulation* — the fact that the articulators for one sound are influenced by those of a surrounding sound. Speech is *co-produced* — an upcoming sound can influence an articulator or set of articulators (an *anticipatory coarticulation*), and a given sound often has leftover influences from a sound that was just made (referred to as a *perseverative coarticulation*). The result is the same; sounds next to each other becoming more similar. Chapter 4 gives more information on anticipatory and perspective coarticulation.

Table 8-1 shows some major varieties of assimilation.

Table 8-1 Assimilation in Action

Example	*Realization (IPA)*	*Explanation*	*Assimilation Type*
ba*d g*uy	[bæg˺ gaɪ]	Place of an alveolar stop assimilates to the place of the following consonant.	*Regressive* (or right to left, anticipatory coarticulation)
captain	[ˈkʰæpm̩]	Place of preceding consonant has an effect on the place of a following one.	*Progressive* (or left to right, perseverative coarticulation)
pan	[pʰæ̃n]	The vowel becomes nasalized before a nasal consonant.	*Similitude* (the phone still sounds like it's in the same category)
sandwich	[ˈsæ̃mɪtʃ]	The /n/ and /w/ are coproduced and fused into a /m/ gesture.	*Coalescence,* also referred to as *fusion* (two sounds merge to form a new segment)

From this table, notice assimilation can proceed in two directions.

- In the first example, "bad guy," a sound segment [g] modifies an earlier sound, which is called *regressive* (or right-to-left) assimilation. You can see a similar direction in the word, "pan," although the process results in a sound just having a slight change that doesn't alter its phonemic status (referred to as *similitude*).
- In contrast, "captain" goes in the opposite direction. The production of [p] affects the place of articulation of the following nasal, [m], a progressive or left-to-right effect. *Progressive* means that a given sound affects the sound following it.
- Finally, "sandwich" illustrates a fusion of two sounds (/n/ and /w/) to result in /m/. This is called *coalescence* because the result of having two distinct phonemes affect each other is a third, different sound.

These examples come from English where harmony cases are local. However, languages such as Turkish and Hungarian have long distance vowel assimilation because these processes cross more than one segment. Refer to the nearby sidebar for a closer look at Hungarian.

Transcribing Hungarian: If you're into hard-core assimilation

Hungarian has a set of front versus back vowels. When a word root contains a back vowel, it must take a suffix with a back vowel. If the word root ends in a front vowel, the suffix must contain a front vowel. Thus, to form the *dative* (indirect object, as in I gave it *to him*) part of speech, Hungarian would form words like this:

Vowel Type	Root	Dative	Gloss
Back vowel	/fal/	/falnak/	"wall"
Front vowel	/kert/	/kertnek/	"garden"

A front vowel triggers assimilation to a front vowel later in the word, and a back vowel triggers a back vowel later in the word. (Similar processes also take place in Turkish, Finnish, and a number of other languages.) For instance, take a gander at these two Hungarian word endings and see if you can spot the regressive assimilation processes at work:

Vowel Type	Root	Dative	Gloss
Back vowel	/had/	/hadnak/	"army"
Front vowel	/hit/	/hitnek/	"belief"

Getting More Different: Dissimilation

Dissimilation is a process where two close sounds become less alike with respect to some property. In dissimilation, sounds march to a different drummer and become less similar. For instance, if a language requires sounds next to each other that are difficult to produce, dissimilation processes come into play so that the final realizations are bold, clean, and producible.

An example is the word "diphthong," which should be pronounced [ˈdɪfθɔ̃ŋ], but is frequently mispronounced [ˈdɪpθɔ̃ŋ]. In fact, many people end up misspelling it as "di*p*thong" for this (mispronunciation) reason.

Because producing a /f/ followed by a /θ/ is difficult (go ahead and try it), two fricatives in a row change to a stop followed by a fricative. Dissimilation isn't quite as common among languages as assimilation.

Putting Stuff In and Out

Processes of *insertion* (also called *epenthesis*) cause a segment not present at the phonemic level to be added. In other words, an unwanted sound gets added to a word.

A common example in English is the insertion of a voiceless stop between a nasal stop and voiceless fricative. Here are some examples:

Example	*Phonemic Level*	*Phonetic Level*
strength	/stɹɛŋθ/	[stɹɛ̃ŋkθ]
hamster	/ˈhæmstəˠ/	[ˈhæ̃mpstəˠ]

Another form of insertion sometimes noted in the language classroom occurs with consonant clusters. Native speakers of languages such as Japanese or Mandarin who don't have consonant clusters (such as pl-, kl-, spr-, or -lk) sometimes insert a vowel between the consonants to make the sounds more like their native phonology. Thus, a Japanese speaker learning English may pronounce the following English words with these epenthetic vowels inserted (in *italics*):

Example	*Phonemic Level*	*Phonetic Level*
spoon	/spun/	[suˈpuːnu]
ski slope	/ski.sloʊp/	[suˈkiːsuˈɹoːpu]

Deletion rules eliminate a sound. An example in English is called h-dropping (or */h/-deletion*). Try and say this sentence, "I sat on his horse." Which of the following two work?

- [aɪ ˈsæſ̃n ɪs ɔrs]
- [aɪ ˈsæt ɔ̃n hɪz hɔrs]

Probably the first is more natural, where /h/ is deleted from "his" and "horse."

Moving Things Around: Metathesis

In *metathesis,* a speaker changes the order of sounds. Basically, one sound is swapped for another. Check out these examples:

Example	*Phonemic Level*	*Phonetic Level*	*Found*
asked	/æskt/	[ækst]	Southern States' dialects and African American Vernacular English (AAVE)
animal	/ˈænɪməl/	[ˈæ̃mĩnəɬ]	Child language

Putting the Rules Together

Some phonological rules depend on others and either set up another rule to operate or deprive them of their chance. The rules in this chapter can all be represented with a basic format:

$A \rightarrow B/C \_\_ (D)$

A becomes B, in the environment after C or before D.

With this format, the following clarify what each letter stands for:

- **A:** The letter on the left side of the arrow is called the *structural description.* This is the sound (at the phonemic level) before anything happens to it.
- **B:** The letter to the right of the arrow is the *structural change.* It's the result of a sound change occurring in a certain environment.
- **C** and **D:** They represent that environment where the sound change occurs.

From the earlier section, "Getting More Alike: Assimilation," I now show the examples in phonological rule format here:

Example	*Structural Description (A)*	*Structural Change (B)*	*Phonological Rule*
ba*d* guy	/bæd gaɪ/	[bæg gaɪ]	Alveolar → + velar/_ velar
pan	/pæn/	[pæ̃n]	Vowel→ + nasal/_ nasal

Instead of writing out lengthy prose, you can use rules to represent phonological processes. For example, for "pan" the rule is that a vowel becomes nasalized in the environment before a nasal.

Consider for a moment how the (tricky) English plurals are pronounced in most words. Although the plural marker -s or -es is used in spelling, it doesn't always result in an [s] pronunciation. Rather, a plural is sometimes pronounced as [s], sometimes as [z], and sometimes as [ɪz], depending on the final sound of the root, as the following examples demonstrate:

Singular	*IPA of Plural Form*	*Suffix*
ra*t*	/ɹæts/	[s]
da*d*	/dædz/	[z]
di*sh*	/ˈdɪʃɪz/	[ɪz]

The plural "s" is a kind of *morpheme*, the smallest meaningful units in language. The study of how *morphemes* show regular sound change is called *morphophonology*. To make the English plural system work, phonologists make two assumptions:

- /z/ is the underlying form of the plural marker.
- Two rules must apply and apply in the correct order.

Table 8-2 specifies these two rules.

Table 8-2 Two Rules of Morphophonology

Rule	*Formula*	*Translation*
Rule No. 1	Insertion: /0/→[ɪ] / [+ sibilant] __ [+ sibilant]	[ɪ] is inserted between two sibilants.
Rule No. 2	Assimilation: /z/ → [-voiced]/ [-voiced, + cons] ___#	[z] becomes devoiced after a voiceless consonant at the end of the word.

In Rule No. 2, the hashmark (#) is an abbreviation for boundary at the end of a word.

You must apply the rules in order for the system to work. If you don't, it bombs. Consider the word "dishes" that ends with [ɪz]:

Singular: [dɪʃ]

Plural: [ˈdɪʃɪz]

Table 8-3 shows correct rules applied in the order. However, the reverse order with Rule No. 2 first doesn't give the right answer. Assimilation changes the /z/ to /s/, then insertion changes the /s/ to /ɪs/, yielding [ˈdɪʃɪs] (incorrect).

Table 8-3 Insertion and Assimilation in Action

First Order (Correct)		*Second Order (Incorrect)*	
Underlying representation	/dɪʃ/ +/z/	Underlying representation	/dɪʃ/ +/z/
Rule No. 1: Insertion	/dɪʃ/ +/ɪz/	Rule No. 2: Assimilation	/dɪʃ /+/s/
Rule No. 2: Assimilation	N/A	Rule No. 1: Insertion	/dɪʃ/ + /ɪs/
Phonetic realization	[ˈdɪʃɪz]	Phonetic realization	[ˈdɪʃ + ɪs]*

Chapter 9

Perusing the Phonological Rules of English

In This Chapter

- Narrowing in on consonant allophones
- Recognizing principled change in vowels
- Getting rule application *just right!*

Phonological rules describe sound processes in language that are naturally understood by speakers and listeners. In order to transcribe well, particularly when completing narrow transcription, it's important to understand these sound processes and describe their output using the correct symbols in the International Phonetic Alphabet (IPA).

Phonological rules take the following form:

Structural description → Structural change /__ (in some environment)

The *structural description* is the condition that the rule applies to. The *structural change* is the result of the rule, occurring in a specific phonetic context. The arrow shows that a given input sound (the structural description) changes or becomes modified in some environment.

A phonological rule can be described in a short description or in a formula. To keep things simple, in this chapter I focus on descriptions upfront to help you understand. I also include a few technical formulas as secondary information. Make sure to check out Chapter 8 for more background about phonology and phonological rules.

There is no set number of phonological rules for any given language. In this chapter, I use 13 phonological rules to capture some of the most important regularities of English phonology. These rules describe *implicit* (naturally understood) processes of a language. The exact numbering doesn't really matter: I group these rules into sections to make them easier to memorize.

As a transcriber, use these rules as a guide for what talkers likely do, but let your ear be the final judge for what you end up transcribing.

Rule No. 1: Stop Consonant Aspiration

A traditional first rule in phonetics is that English voiceless stops, which are /p/, /t/, and /k/, become aspirated when stressed and syllable initial (at the beginning of a syllable). This rule captures that fact that the phoneme /t/ is represented by the aspirated allophone [t^h] under these specific conditions.

Each phonological rule usually has an IPA diacritic or symbol involved. As a result, I list relevant diacritics and symbols following each rule. I also provide some examples, and I encourage you to generate your own. The diacritic for Rule No. 1 is [^{h}]. Here are some examples:

*p*eace [$p^h$is]

a*tt*ire [ə$^{\text{ˈ}}t^h$aɪɹ]

*k*iss [k^hɪs]

This rule captures one of the essential properties of English phonology. Try saying each word while holding your hand under your mouth (near your bottom lip) and you should feel a puff of air that is the aspiration of the [p^h], [t^h], and [k^h].

Monosyllabic words, those words that have just one syllable, such as "peace" and "kiss," are easy. However, in *polysyllabic words,* words with multiple syllables, things get a bit more complicated. Aspiration is stronger in stressed syllables than unstressed (see Chapter 6 for further discussion), which means in polysyllabic words the aspiration rule applies chiefly to stressed syllables. Otherwise, the /p/, /t/, and /k/ consonants are released, but not stressed. Here are some examples of polysyllabic words:

catapult [$^{\text{ˈ}}k^h$ærəpəɬt]

repulsive [ɹə$^{\text{ˈ}}p^h$ʌɬsɪv]

Try the aspiration test, feeling for an air puff, when saying "catapult" and "repulsive." Aspiration is on the initial stop in "*c*atapult" because the [k^h] is syllable-initial and stressed. However, even though the [p] in "cata*p*ult" is syllable initial, it isn't aspirated. It's only released. In "re*p*ulsive," the [p] is aspirated because it's syllable-initial and stressed (even though it's not word initial).

Aspiration for English /p/, /t/, and /k/ generally isn't as strong when word-initial than, for example, when following another word. *Word initial* means at the beginning of a word, so the [p^h] in "*p*ie" generally has less aspiration than the [p^h] in "the *p*ie." For this reason, you may see different conventions used by phoneticians when marking aspiration in narrow transcriptions at the beginning of words. Some mark it and others don't. In this book, I mark aspiration at the beginning of a word, according to Rule No. 1.

Table 9-1 includes some practice items containing /p/, /t/, and /k/. Mark the aspiration using narrow transcription in column three. I have done the first one for you. Ready?

Table 9-1 Stop Consonant Aspiration Practice

Example Word	*Broad*	*Narrow*
appear	/əˈpɪɹ/	[əˈpʰɪɹ]
khaki	/ˈkæki/	
uncouth	/ənˈkuθ/	

The answers are as follows:

- **khaki:** You should only have marked the initial [k^h] of "khaki" as aspirated because that "k" is stressed and syllable initial. The second [k] is released but not aspirated.
- **uncouth:** The [k] is aspirated because it's stressed and syllable initial, even though it's the final syllable in the word.

If you're more into formulas, you can write Rule No. 1 as:

C [+stop, –voice] → [+aspiration]/ #___ [+ syllable, + stress],
(where # = boundary)

Here's how to read this formula. "A consonant (that is a stop and is voiceless) becomes aspirated in the environment at the beginning of a stressed syllable." Or more simply, stop consonants are aspirated in stressed syllable-initial position.

Rule No. 2: Aspiration Blocked by /s/

Another rule of phonetics is voiceless stops become unaspirated after /s/ at the beginning of a syllable. Because English has many consonant clusters (groups of consonants in a row, such as [spɹ] and [sk]), some phonologists consider this an important rule to remember. Others note that it overlaps with Rule No. 1. I emphasize this rule because it shows the importance of rule interaction. ***Note:*** There really is no diacritic or symbol for this rule because a feature is being blocked, not added.

English syllable-initial, s-containing consonant clusters (sp-, st-, and sk-) all share something in common: the production of the /s/ blocks the following stop consonant from having much aspiration. Try the following examples of minimal pairs, putting your hand near your mouth for the aspiration test.

*p*ill [p^hɪɬ]	*sp*ill [spɪɬ]
*t*ill [t^hɪɬ]	*st*ill [stɪɬ]
*k*ale [k^heɬ]	*sc*ale [skeɬ]

Notice that this rule would *not* apply in words such as "wa*sp*," "wri*st,"* or "fla*sk*," where the s-clusters occur at the end of a syllable. In such cases, the structural description isn't met and the rule isn't relevant. In words such as "whi*sp*er" (s-cluster in the medial position), the rule does apply because the stop comes after /s/ and at the beginning of a syllable. Try the aspiration test for "whisper" and see for yourself! No aspiration should be notable on the [p].

If you enjoy formulas, you can write phonological Rule No. 2 as follows:

C [+stop, -voice] → [-aspiration]/ #s__ [+ syllable]

A rough description of this rule would be: "Consonants that are stops and voiceless don't become aspirated when following an /s/ at the beginning of a syllable." Or more simply, voiceless stops become unaspirated after syllable-initial /s/.

Rule No. 3: Approximant Partial Devoicing

Devoicing rules are a rather depressing thing for phonetics teachers to talk about because it reminds them that life can get really complicated. When someone first starts to study phonetics, voicing is a comfortable, solid binary feature. A *phone* (speech sound) is defined as voiced or voiceless, end of

story. However, then a dirty little secret comes out: Under certain conditions, some sounds may become partially *devoiced* (spoken with less buzzing of the vocal folds) because of biomechanical and timing reasons.

If you're making an aspirated stop such as [p^h] in "*p*ay," the aspiration will affect the following approximant, such as if you say "p*r*ay" or "*pl*ay." For "pray" or "play" the vocal folds won't have time to fully buzz for the [ɹ] and [l], resulting in partial devoicing.

The diacritic for partial devoicing is a small circle placed under the sound, [◌̥]. Here are some examples for Rule No. 3:

*pr*ay [p^h̥ɹe]

*cl*ass [k^h̥læs]

*tw*ice [t^h̥waɪs]

*cu*te [k^h̥jut]

You can get a good sense of how this works by placing your hand over your voice box to feel the buzzing while you say the following examples of minimal pairs.

*r*ay [ɹe] — p*r*ay [p^h̥ɹe]

*l*ass [læs] — c*l*ass [k^h̥læs]

*w*eak [wik] — t*w*eak [t^h̥wik]

you [ju] — *cu*e [k^h̥ju]

You should feel a longer period of buzzing for the (italicized) approximants in the list when they aren't preceded by a [p^h], [t^h], or [k^h].

You place a small circle beneath the symbol as a diacritic indicating partial devoicing (unless the font is too cluttered by a downward-going symbol, such as /j/, in which case the symbol can be placed above the character).

If you're more into formulas, you can write phonological Rule No. 3 as:

C [+approximant] → [–voice]/ C [+ stop, +aspiration] __

This formula reads "consonants that are approximants become partially devoiced in the environment following consonants that are stops and are aspirated." Or more practically, "approximants become (partially) devoiced after aspirated stops."

Rule No. 4: Stops Are Unreleased before Stops

A *release burst* occurs when a stop consonant closure is opened, producing a sudden impulse that is usually audible. In aspirated stops at the beginning of a syllable (like [p^h] in "*p*et" [$p^h\varepsilon t$]), the vocal folds are apart, and there's aspiration (breathy, voiceless airflow) after the release of the stop. Try it and you can feel the aspiration on your hand or watch a candle blow out. English syllable-initial voiced stops (as in *b*et [bɛt]), also have a burst, but without aspiration and with a shorter voice onset time (VOT, see Chapter 15 for more info). This release burst energy is weaker but is usually audible.

Release bursts may not be audible in other situations. These situations are referred to as *no audible release.* In syllable-final position, stop consonants can be optionally released. Try saying "tap/tab" in two speaking conditions:

- Quickly and casually: In casual speech, people usually produce no audible release for a syllable-final stop.
- Carefully, as if you were addressing a large audience that could barely make out what you were saying: More formal speech can override this no audible release condition. In formal speech, release characteristics are often emphasized for clarity or style.

When a person produces two stops in a row, the release characteristics become poorly audible. For instance, when saying "risked" [ɹɪsk˺t], just as the vocal tract is being configured for the release burst of the /k/, the tongue is also making closure for /t/, effectively cancelling out any sound of a released /k/. The diacritic for this rule is [˺]. Some examples are

risked [ɹɪsk˺t]

bumped [bʌm˺pt]

To see how hard it is to produce the word "risk" (with release) followed by a /t/, try these four steps:

1. **Produce "risk" casually with no audible release.**

   risk [ɹɪsk]: No special diacritic is needed to mark lack of audible release.

2. **Add the final /t/ to "risk."**

   risked [ɹɪsk˺t]: This is the normal output of Rule 4.

3. **Produce "risk" with a full release.**

   risk [ɹɪsk^h]: The aspiration diacritic is used only if the final release is strong enough to warrant it.

4. **Add the final /t/ again.**

   risked [ɹɪskʰt]: Argh! This won't sound natural.

Rule No. 5: Glottal Stopping at Word Beginning

A rather surprising use of the glottal stop in English occurs before vowels at the beginning of a word or phrase. Unless you ease into an utterance (making some kind of ultra-calm announcement to zoned-out meditators at a Yoga retreat), you probably precede a vowel with a glottal stop. The IPA character that you need to remember for this rule is the glottal stop ([ʔ]).

Here are some examples. Try them and pay attention to whether your glottis is open or closed.

eye [ʔaɪ]

eaten [ˈʔiʔn̩]

Some phoneticians consider this rule in transcriptions and some don't. I use word-initial glottal stopping in the optional transcriptions listed in the audio-visual materials located at `www.dummies.com/go/phoneticsfd`.

Rule No. 6: Glottal Stopping at Word End

Voiceless stops are preceded by glottal stops after a vowel and at the end of a word. This rule also applies to word-final voiceless affricates. The IPA symbol involved in this rule is the glottal stop [ʔ]. Some examples include

steep [stiʔp]

pitch [pʰɪʔtʃ]

This rule is a use of a glottal stop that many English speakers don't believe at first, but eventually they'll accept. Before syllable-final /p/, /t/, /k/, or /tʃ/, many speakers of English restrict the flow of air at the glottis before getting to the stop itself (or at the same time as realizing the stop). Such timing doesn't occur

if the final stop is voiced. Try these following words and see if you pronounce the voiceless stops in such a manner:

ri*p* [ɹɪʔp]

ri*ch* [ɹɪʔʧ]

rib [ɹɪb]

ridge [ɹɪʤ]

Whether glottal stop in this pre-consonantal position is transcribed or not is generally up to the discretion of the phonetician. Some capture this detail in narrow transcription and others don't. I provide this detail (as alternate transcriptions) in the audiovisual materials (located at `www.dummies.com/go/phoneticsfd`).

Rule No. 7: Glottal Stopping before Nasals

Here is another rule that describes the distribution of glottal stop: "Voiceless alveolar stops become glottal stops before a nasal in the same word." In other words, this rule captures the fact that /t/ and /d/ become [ʔ] in certain environments.

The symbol for this rule is the glottal stop [ʔ]. Say these words and think about what they all have in common:

ea*t*en [ˈiʔn̩]

wri*tt*en [ˈɹɪʔn̩]

bi*tt*en[ˈbɪʔn̩]

ro*tt*en[ˈɹɑʔn̩]

ki*tt*en[ˈkɪʔn̩]

glu*tt*on[ˈglʌʔn̩]

If you speak North American English, you'll almost certainly pronounce the medial /t/ phoneme as glottal stop [ʔ], followed by a syllabic nasal, indicated by placing a small line below the [n̩], described in Rule No. 9, explained later in this chapter.

Notice that none of these word examples involve an aspirated medial /t/ phoneme ([t^h]). Also, the stress pattern is *trochaic* (which means the syllable's stress is strong, then weak, sounding *loud*-soft (as in "*rif*le" "*doub*le", and "*ti*ger").

Rule No. 8: Tapping Your Alveolars

Alveolar stops (/t/ or /d/) become a voiced tap between a stressed vowel and an unstressed vowel. A *tap* (also called *flap* by some phoneticians, see Chapter 6) is a rapid articulation in which one articulator makes contact with another. Unlike a stop, there's not enough time to build up a release burst.

This rule involves the IPA symbol [ɾ], an English *allophone*. That is, a tap can't stand by itself anywhere in the language to change meaning. In English, a tap only occurs in certain environments, as specified by phonological rules. Here are some examples:

glottal [ˈglɑɾt̬]

Betty [ˈbɛɾi]

daddy [ˈdæɾi]

The stress patterns of the words involved are trochaic, like the cases in Rule No. 7. If there were someone named "Beh *Tee*," for example, this tapping rule wouldn't work! In such a case, the alveolar stop would instead be aspirated: [bɛˈtʰi]. Some speakers of North American English may produce medial /d/ as more of a voiced stop than a tap, thus pronouncing "daddy" as [ˈdædi].

Rule No. 9: Nasals Becoming Syllabic

This rule states that nasals become syllabic at the end of a word and after an *obstruent* (such as fricatives, stops, and affricates). In broad transcription, words ending with (spelled) "-en" and "-em" are represented using the IPA symbols /ən/ and /əm/. However, broad transcription doesn't capture all the possibilities for these sounds. The diacritic for this rule is a small vertical line placed under the nasal consonant [ˌ].

For instance, in the word "button," you usually don't include much [ə] vowel quality in the final syllable. Instead, you make a nasal release by lowering the soft palate, rather than the tongue, which results in a pure "n" that stands by itself as a syllable. Here are the broad narrow transcriptions for "button."

Broad: /ˈbʌtən/

Narrow: [ˈbʌʔn̩]

Try just holding out a long "n". You can do this without any [ə] vowel quality. To transcribe a syllabic nasal "n," you place a small vertical line under the character, like this: [n̩]. You transcribe syllabic "m" like this: [m̩].

Here are some examples in a GAE accent, narrowly transcribed:

written [ˈɹɪʔn̩]

bottom [ˈbɑɾm̩]

Rule No. 10: Liquids Become Syllabic

This rule is very similar to Rule No. 9; however, it applies to sounds that are typically spelled with "-er" and "-el". In certain environments, sounds that are broadly transcribed /ɚ/ or /əl/ are in fact produced syllabically, [ɻ̩] and [ɬ̩]. This rule has the same diacritic as Rule No. 9, a combining small vertical bar under the consonant [ˌ].

The following examples compare broad and narrow transcriptions for words containing liquid consonants (/ɹ/ and /l/):

Word Example	***Broad IPA***	***Narrow IPA***
couple	/ˈkʌpəl/	[ˈkʰʌpɬ̩]
writer	/ˈɹaɪtɚ/	[ˈɹaɪɾɻ̩]

The word "couple" has a lateral release of the plosive. Say the word and pay attention to the final syllable; you'll probably find not much [ə] vowel quality.

The case for (spelled) "-er" is more ambiguous: Some phoneticians use syllabic "r" ([ɻ̩]) in narrow transcription for words like "writer." Others point out that syllabic "r" is equivalent to [ɚ] in most cases, and tend to use this syllabic diacritic less. I use syllabic "r" in narrow transcription, following Rule No. 8.

In these words, like the nasal examples in Rule No. 9, the syllabified liquids occur in the unstressed syllable of trochaic (*loud*-soft) word patterns.

Rule No. 11: Alveolars Become Dentalized before Dentals

This is an *assimilation* rule, where one sound becomes more like its neighbor. The main influencing sounds are the interdentals /θ/ and /ð/, which can influence a number of alveolars (/n/, /l/, /t/, /d/, /s/, /ɹ/, and /z/). The dental fricatives in English (/θ/ and /ð/) are also called *interdentals* because they involve airflow between the upper and lower teeth.

The diacritic associated with Rule No. 11 is a small square bracket, that looks like a staple, placed under the consonant: [̪].

When an alveolar consonant is produced before a *dental* (sound produced against the teeth), the alveolar is produced more forward than usual. This is called being *dentalized* because the affected sound is now made closer to the teeth.

Try these minimal pair examples, paying attention to where your tongue tip is at the end of each alveolar (italicized) sound.

te*n* [tʰɛn]	te*n*th [tʰɛn̪θ]
fi*ll* [fɪɬ]	fi*l*th [fɪɬ̪θ]
no*r*[ˈnɔɹ]	no*r*th [ˈnɔɹ̪θ]

Rule No. 12: Laterals Become Velarized

This rule refers to the English lateral ("l" consonant) becoming *dark* (velarized) in certain environments, otherwise remaining *light* (clear, or alveolar). Specifically, laterals become velarized after a vowel and before a consonant or at the end of a word.

If you sing *la-la-la,* you can remember that light (or clear) "l" comes at the beginning of syllables, while dark "l" is at the end. Another way of distinguishing dark from light "l" is to use your ear: The sound of "l" at the end of the word "little" (syllable final) sounds much lower than the "l" at the beginning (syllable initial).

The diacritic used to denote velarization is a tilde placed in the middle of an IPA character. For instance, velarized "l" is written as [ɬ]. A couple examples of this rule are

waffle [ˈwɑfɬ]

silk [sɪɬk]

Rule No. 13: Vowels Become Nasalized before Nasals

If you happen to be a speaker of Portuguese, you'll have fairly precise control of nasality in vowels because this serves meaning in your language. This is because nasality is phonemic in Portuguese; it matters to the listener. However, in English nasality spreads from a consonant onto the vowel in front of it. As such, there is much variation from talker to talker: Some people partially nasal the vowel and others nasalize it entirely. The amount doesn't matter that much to the listener.

The diacritic for nasalization is a tilde placed over a vowel symbol is [˜]. Some examples of this rule are

seem [sĩm]

soon [sũn]

As a transcriber your job on this one is easy. Every time you see a vowel in front of a nasal, that vowel is nasalized. This isn't indicated in broad transcription, but is usually marked in narrow. Table 9-2 gives you some examples.

Table 9-2 Examples of Nasalized Vowels

Example Word	*Broad*	*Narrow*
banana	/bəˈnænə/	[bə̃ˈnæ̃nə]
incomplete	/ɪnkəmˈplit/	[ĩnˈkə̃mˈ̩pʰl̥iʔt]
camping	/ˈkæmpɪŋ/	[ˈkʰæ̃mˈ̩pm̃ŋ]

In addition to noting how the nasality rule (Rule No. 13) operated on these words, can you also see how a consonant glottalizing rule (Rule No. 5) and a

stop release rule (Rule No. 4) applied? How about aspiration (Rule No. 1) and approximant partial devoicing (Rule No. 3)?

Applying the Rules

It's one thing to know these rules in this chapter; it's another to apply them. Beginning transcribers sometimes have trouble using the rules of English phonology to complete narrow transcriptions. In this section, I show you the most common errors made and provide a quiz to get you started on the right track.

Are you a phonological rule over-applier, under-applier, or "just right"? Take this simple test. Some people are phonological *rule under appliers*. Due to extreme caution (or perhaps just due to confusion) these folks tend to not apply rules where needed. In contrast, others take a *What the heck!* approach and plaster rules all over the place, even when such rules could not conceivable apply. For instance, aspiration may be placed on fricatives, nasalization over stops, and so forth.

Table 9-3 shows some examples of these three types of transcribers. Look to see where you fall.

Table 9-3 Three Degrees of Transcribers

Example Word		*Under-applied*	*Over-applied*	*"Just right"*
pants	/pænts/	[pænts]	[p^hãn˺ʔs̩]	[p^hãn˺ts]
pack rats	/ˈpækɹæts/	/pækɹæts/	[ˈp^hæk^hɹæt^hs]	[ˈp^hækɹæts]

In "pants" (American English accent), the syllable-initial /p/ would ordinarily be aspirated and the nasal vowel would be nasalized (as shown in the "just right" column). Here, an under-applier might note nothing, while the over-applier throws in a gratuitous syllabic symbol under the [s], which would make "pants" a two-syllable word.

In "pack rats," the under-applier again misses all rules. In this case, stress assignment is also missed. The over-applier liberally sprinkles aspirations everywhere, even when they don't apply. Just because voiceless stops can be aspirated doesn't mean they *are* (the rule notes this occurs only in syllable initial position).

To avoid such boo-boos, remember that you don't need to use all diacritics all the time. Only use them if they're absolutely needed. Every diacritic counts. Phoneticians are picky. Ready for a quick quiz?

Which of the following narrow transcriptions would apply to the following broad transcription of "crunch," (/kɹʌntʃ/) as produced by someone from North America?

a. [k^{h}ə̹ɹʌ̃ʔn̚ʧ]

b. [k^{h}ɹ̥ʌ̃ntʃ]

c. [kɹʌntʃ]

d. All of the above are correct

The correct answer is *b*.

If you answered *a,* you over-applied the rules. If you answered *c*, you under-applied. Answer *d* is incorrect, because *a* and *c* are highly unlikely narrow transcriptions of /kɹʌntʃ/.

Chapter 10

Grasping the Melody of Language

In This Chapter

- Using juncture for different speaking styles and rates
- Exploring the syllable and stress assignment
- Patching with sonority and prominence measures

Transcribing is more than just getting the vowels and consonants down on paper. You need that extra zest! For instance, you should be able to describe how phonemes and syllables join together, a property called *juncture*. A phonetician must be able to hear and describe the melody of language, focusing on patterns meaningful for language. This important sound aspect, called *prosody,* gives speech its zing and is described with a number of specialized terms. This chapter gives you the tools to handle bigger chunks of language, so that you can master description of the melody of language.

Joining Words with Juncture

Unless you're a lifeless android (or have simply had a very bad night), you probably don't say things such as "Hel-lo-how-are-you-to-day?" That is, people don't often speak one word (or syllable) at a time. Instead, speech sounds naturally flow together. *Juncture* is the degree to which words and syllables are connected in a language. These sections explain some characteristics of juncture and help you transcribe it.

Knowing what affects juncture

A number of factors can affect juncture, including the following:

- **Some factors are language-specific.** Some languages (such as Hawaiian) break things up and have relatively little carryover between syllables, while other languages (such as French) allow sounds to be run together.

In French, the process of sounds blending into each other is called *liaison,* in which sounds change across word boundaries. Check out these two examples:

Language	*Spelling*	*IPA*	*Translation*
Hawaiian	humu humu nuku nuku apu a u a	/hu.mu.ˌhu.mu.ˌnu. ku.ˌnu.ku.ˌaː.puˈa.ʔa/	Small reef triggerfish
French	les amis	/le‿zaˈmi/	the friends

In these examples, the syllables of Hawaiian have little effect on each other, whereas the French has *resyllabification* (the shift of a syllable boundary) and a voicing of an underlying /s/ sound — a clear example of adjacent sounds affecting each other.

- **Other factors are more personal.** They include speaking formality and rate. Think about how your speech changes when you formally address a group versus talking casually with your friends. In a formal setting, you usually use more polite forms of address (*sir* and *madam*), fancier terms for things (*restroom* or *public convenience* instead of *john* or *loo*), and frillier sentence constructions (*Would you kindly pass the hors d'oeuvre please?* instead of *Yo. The cheese, please?*).

  In informal speech, talkers usually have less precise boundaries than in formal speech. This register change often interacts with *rate,* because rapid speech often causes people to *undershoot* articulatory positions (not reach full articulatory positions). The result can be *vowel centralization* (sounds taking on more of an [ə]-like quality), *de-diphthongization* (diphthongs becoming monophthongs), changes in *consonant quality* (such as the tongue moving less completely to make speech sounds), and changes in *juncture boundaries* (including one boundary shifting into another).

  Check out these examples from American and British English:

Language	*Example*	*Formal (Slow Rate)*	*Informal (Fast Rate)*
American English	How are you?	/haʊ ˈɑɹ ju/	[həˈwɑɹjə]
British Standard English	Nice, isn't it?	/naɪs ˈɪznt ɪt/	[ˈnaɪsɪnʔɪ]

Changes in register and style clearly affect juncture (how speech sounds are connected in terms of pauses or gaps). Some phoneticians refer to juncture as *oral punctuation* because it acts somewhat like the commas and periods in written language.

Transcribing juncture

You can transcribe juncture in a couple different ways. They are as follows:

- **Close juncture:** This default way of transcribing shows that sounds are close together by placing IPA symbols close together in transcription from phoneme to phoneme. An example is "Have a nice day!" /hævə naɪs ˈdeɪ/.
- **Open juncture:** You use *open juncture* (also referred to as *plus juncture*) symbols when you need to emphasize gaps separating sounds. Consider these two expressions:

  "Have a nice day!" /ˈhævə + naɪs ˈdeɪ/

  "Have an ice day!" /ˈhævən + aɪs deɪ/

  Many speakers would probably produce this second example ("Have an ice day") with a glottal stop before the vowel of ice, as a way of marking the gap between the words "an" and "ice." To distinguish these two expressions, the exact placement of the gap between the /ə/ and /n/ is critical. Therefore, open juncture symbols are helpful.

Phoneticians use different conventions for juncture between words. Depending on the speaking style, some phoneticians place a content word (such as the verb "have" in the preceding examples) next to an adjacent function word (such as the determiner "a"), resulting in /ˈhævə/. Doing so tells the reader there is no pause between these sounds. Other transcribers indicate such juncture with a tie-bar at the bottom of the two words: (/ˈhæv‿ə/).

The flow of spoken language doesn't necessarily follow the grammatical patterns you learned in English class. Talkers can run-on or hesitate during speech for many reasons. Consider the sentence, "I went to the store." This sentence can be produced with many different juncture patterns, such as

I . . . went to the store.

I went . . . to the store.

I went to . . . the store.

I went to the . . . store.

And so on. You get the idea. Transcribing all the potential variations in the exact same way wouldn't make sense. What's important is showing where all the gaps take place. Many phoneticians use the IPA pipe symbol ([|]), which technically indicates a *minor foot,* a prosodic unit that acts like a comma (I describe it in greater detail in Chapter 11). However, many transcribers also use this symbol to represent a short pause, whereas they use a double bar

([‖]) to represent a long pause, such as at the end of a sentence. Here are some examples:

/aɪ |wɛn tə ðə stɔɹ‖/

/aɪ wɛnt | tə ðə stɔɹ‖/

If you use these symbols in this manner, be sure to indicate it in notes to your transcription. A good general principle to follow is to employ juncture and timing information only when needed. For instance, the hash mark (#) is a linguistic symbol that means a boundary, such as the end of a word. I have seen older phonetic transcriptions with a hash mark placed between every word. These ended up looking as if a psychotic chicken used the transcription to practice the Rhumba. Keep your transcriptions tailored to your needs, with just the amount of detail your applications require.

Emphasizing Your Syllables

A syllable is something everyone knows intuitively, but can drive phoneticians nuts trying to pin down precisely. By definition, a *syllable* is a unit of spoken language consisting of a single uninterrupted sound formed by a vowel, diphthong, or syllabic consonant, with other sounds preceding or following it. Phoneticians don't see the definition so cut and dry.

Phoneticians consider a *syllable* an essential unit of speech production. It's a unit with a center having a louder portion (made with more air flow) and optional ends having quieter portions (made with less air flow). Phoneticians agree on descriptive components of an English syllable, as shown in Figure 10-1.

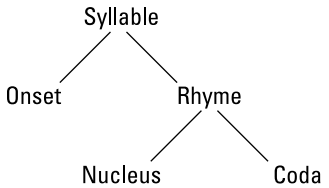


Illustration by Wiley, Composition Services Graphics

Figure 10-1: Parts of an English syllable.

From Figure 10-1, you can see that an English syllable (often represented by the symbol *sigma* [σ]), consists of an optional *onset* (beginning) and a *rhyme* (main part). The rhyming part consists of the vowel and any consonants that come after it. The vowels in a rhyme sound alike. At a finer level of description, the rhyme is divided into the *nucleus* (the vowel part) and the *coda* (tail or end) where the final consonants are. From this figure, you can take a word like "cat" and identify the different parts of the syllable. For "cat" (/kæt/), the /k/ is the onset, /æ/ is the nucleus, and the /t/ is the coda.

This is why this type of poem rhymes:

> Roses are red, violets are bl*ue*. . . .
>
> blah blah blah *blah,* blah *blah* blah blah . . . y*ou.*

Languages vary considerably with which kinds of onsets and codas are allowed. Table 10-1 shows some samples of syllable types permissible for English.

Table 10-1 Sample Syllable Types in English

Example	*IPA*	*Syllable Type*
eye	/aɪ/	V
hi	/haɪ/	CV
height	/haɪt/	CVC
slight	/slaɪt/	CCVC
sliced	/slaɪst/	CCVCC
sprints	/spɹɪnts/	CCCVCCC

The last column lists a common abbreviation for each syllable type, where "C" represents a consonant and "V" represents a vowel or diphthong. For instance, "eye" is a single diphthong and thus has the syllable structure "V." At the bottom of the table, "sprints" consists of a vowel preceded and followed by three consonants, having the structure "CCCVCCC."

Strings of consonants next to each other are called *consonant clusters* (or *blends*). Each language has its own rules for consonant cluster formation. The permissible types of consonants clusters in English are, well, rather odd. Figure 10-2 shows some of the English initial consonant clusters in a chart created by the famous Danish linguist, Eli Fisher-Jørgensen.

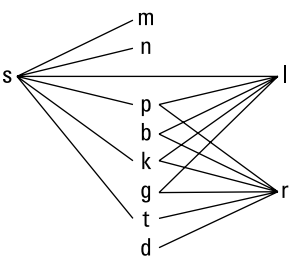


Illustration by Wiley, Composition Services Graphics

Figure 10-2: Some English syllable-initial consonant clusters.

Notice the *phonotactic* (permissible sound combination) constraints at work in Figure 10-2. It's possible to have *sm-* and *sn-* word beginnings, but not *sd-*, *sb-*, or *sg-*. There can be an *spl-* cluster, but not a *ps-* or *psl-* cluster.

Stressing Stress

Nothing makes a person stand out as a foreign speaker more than placing stress on the wrong syl*lab*le. In order to effectively teach English as a second language, transcribe patient notes for speech language pathology purposes, or work with foreign accent reduction, you need to know how and where English stress is assigned. This, in turn, requires an understanding of phonetic stress at the physiologic and acoustic levels.

Stress is a property of English that's signaled by a syllable being louder, longer, and higher than its neighbors. It's a *suprasegmental* property (which means that it extends beyond the individual consonant or vowel). *Louder, longer and higher* are perceptual properties, that is, in the ear of the beholder. For a syllable to be perceived as stressed, physical attributes must be physically changed. For now, this table describes what a talker does to produce each of these speech properties (*articulatory*), what the acoustic property is called (*acoustic change*), and how it's heard (*perceptual impression*). Check out Chapter 12 for more in-depth information.

To understand Table 10-2 and get a sense of how louder, longer, and higher works, say a polysyllabic word correctly and then say it incorrectly. Say "*syl*lable" correctly, with stress on the initial syllable. Next, incorrectly place the stress on the second to last syllable (also called the *penultimate*, or *penult*), as in "syl*la*ble." Finally, place stress on the final syllable, or ultima, "sylla*ble*."

Table 10-2 Physical, Acoustic, and Perceptual Markers of Stress in English

Articulatory	*Acoustic Change*	*Perceptual Impression*
Increased airflow, greater intensity of vocal fold vibration	The amplitude increases	Louder
Increased duration of vocal and consonantal gestures	The duration increases	Longer sound ("length")
Higher rate of vocal fold vibration	The fundamental frequency increases	Higher pitch

In each case (whether you're correctly or incorrectly pronouncing it), the stressed syllable should sound as if someone cranked up the volume. The following sections tell you more about how stress operates at the word, phrase, and sentence level in English.

Eyeing the predictable cases

Stress serves four important roles in English. They are as follows:

- **Lexical (word level):** When you learn an English word, you learn its stress. This is because stress plays a *lexical* (word specific) role in English: it's assigned as part of the English vocabulary. For example, syllable is pronounced /ˈsɪləbəl/, not /sɪˈlʌbəl/ or /sɪləˈbʌl/.
- **Noun/verb pairs:** In English, stress also describes different functions of words. Try saying these noun-verb pairs, and listen how stress alteration makes a difference (the stressed syllables are italicized):

Spelling	*Part of Speech*	*IPA*
(to) re*cord*	Verb	[ɹəˈkʰɔɹd]
(a) *rec*ord	Noun	[ˈɹɛkɚd]
(to) re*bel*	Verb	[ɹəˈbɛɬ]
(a) *re*bel	Noun	[ˈɹɛbɬ]

  These stress contrasts are common in stress-timed languages, such as English and Dutch (whereas tone languages, such as Vietnamese, may distinguish word meaning by contrasts in pitch level or pitch contour on a given syllable).

- **Compounding:** With compounding, two or more words come together to form a new meaning, and more stress is given to the first than the second. For example, the words "black" and "board" create "blackboard" /ˈblækbɔɹd/.

  Also, the juncture is closer than a corresponding adjective + noun construction. For example, if you pronounce the following pairs, you'll notice a longer pause between the words in the first example (the English column) than between the words in the second example (the IPA column).

Grammatical Role	*English*	*IPA*
Adjective + noun	a black board	/ə blæk ˈbɔɹd/
Compound noun	a blackboard	/ə ˈblækbɔɹd/

- **Emphasis in phrases and sentences:** Also known as *focus,* this is a pointer-like function that draws attention to a part of a phrase or sentence. By making a certain syllable's stress louder, longer, and higher, the talker subtly changes the meaning. It's as if the utterance answers a different question. For example:

  Dylan sings better than Caruso. (Who sings better than Caruso?)

  Dylan *sings* better than Caruso. (What does Dylan do better than Caruso?)

  Dylan sings better than *Caruso.* (Who does Dylan sing better than?)

  People handle this kind of subtlety every day without much problem. However, just think how difficult it is to get computers to understand this type of complexity.

Identifying the shifty cases

For the most part, English stress remains fairly consistent. However, some cases realign and readjust. You may think of it as a musical score having to be switched around here and there to keep with the rhythm. These adjustments, called *stress-shift*, are a quirky part of English phonology.

Stress realigns itself in a manner to preserve the up-and-down (rhythmic) patterns of English. If syllables happen to combine such that two stressed syllables butt up against each other, one flips away so that there is some breathing room. Think of it like two magnets with positive and negative ends: put two positives together and one flips around so that it's *positive/negative/positive/negative* again.

Some English words take primary stress on different syllables, based on the context. For example, you can pronounce the word "clarinet" with initial stress, such as /ˈklɛɹɪnɛt/ or with final stress, as in /klɛɹɪˈnɛt/, depending on the stress of the word that comes next. Try this test:

1. **Say "Clari*net mus*ic" three times.**

   Doing so sounds a bit awkward, right? It should have been more difficult because two stressed syllables had to butt up against each other.

2. **Say "*Clar*inet *mus*ic" three times.**

   You should notice that this second pattern flows more naturally because it permits the usual English stress patterns (strong/weak/strong/weak) to persist.

Sticking to the Rhythm

Another way an English speaker can show adeptness with the language is having the ability to use English *sentence rhythm patterns*, where greater

stresses occur at rhythmic intervals, depending on talking speed. To get a sense of these layered rhythms, consider these initially stressed polysyllabic words: "really," "loony," "poodle," "swallowed," "fifty," "plastic," and "noodles."

When you put them together in a sentence, they form:

> The really loony poodle swallowed fifty plastic noodles.

Although speaking this sentence is possible in many fashions, a typical way people produce it is something like this:

> The *real*ly loony *poo*dle swallowed *fift*y plastic *noo*dles.

That is, regularly spaced, strongly stressed syllables (italicized) are interspersed with words that still retain their primary stress (such as "loony"), yet they're relatively deemphasized in sentential context. This kind of timing is rhythmic and can reach high levels in art forms like vocal jazz (or perhaps, rap). Chapter 11 discusses ways you can transcribe this kind of information.

Tuning Up with Intonation

In phonetics, *sentence-level intonation* refers to the melodic patterns over a phrase or sentence that can change meaning. For instance, rising or falling melodic patterns that change a statement to a question, or vice-versa. Intonation is quite different from *tone*, which is the phoneme-level pitch differences that affect word meaning in languages such as Mandarin, Hausa, or Vietnamese (see Chapter 18). English really has no tone. The following sections take a closer look at the three patterns of sentence-level intonation that you find in English.

Making simple declaratives

A basic pattern of English intonation is the simple *declarative* sentence, which is a statement used to convey information. A couple examples are "The sky is blue" or "I have a red pencil box."

Think of this pattern as the plain gray sweater of the phonetic wardrobe. A bit dull, perhaps, but it's necessary. When you're simply stating something, the chances are your intonation is falling. That is, you start high and end low.

Falling intonation seems to be a universal pattern, perhaps due to the fact that it takes energy to sustain the thoracic pressure needed to keep the voice box (larynx) buzzing. As a person talks, the air pressure drops and the amount of buzzing tends to drop, causing the perceived pitch to fall, as well.

Answering yes-no questions

The second pattern of sentences is called the "yes/no question." When you're asking a question that has a yes or no answer, you probably have rising intonation. This means you start low and end high.

Try producing the same sentences that I introduce in the previous section, but instead of falling in pitch as you speak, have your voice rise from low to high.

You probably noticed these English statements ("The sky is blue?") have now turned into questions. Specifically, they're questions that can be answered with *yes* or *no* answers. This rising pitch pattern for questions is fairly common among the world's languages. For instance, French forms most questions in this manner. ***Note:*** Some languages don't use intonation at all to form a question. For instance, Japanese forms questions by simply sticking the particle /ka/ at the end of a sentence.

Focusing on "Wh" questions

The third pattern of sentences include English questions with the *Wh questions*, including "who," "what," "when," "where," "why," and "how," (which are produced with falling pitch, rather than rising). Try a few, while determining whether your voice goes up or down:

Who told you that?

What did he say?

When did he tell you?

Where will they take you?

Why are you going?

How much will it cost?

Your intonation likely goes down over the course of these utterances. Try this for yourself. Say the preceding sentences to see whether your intonation goes down.

Showing Your Emotion in Speech

When someone talks, part of the melody serves a language purpose, and part serves an emotional purpose. When you're transcribing speech, you need to understand emotional prosody because it can interact in complex ways with the linguistic functions of prosody. In fact, people can show many emotions in speech, including joy, disgust, anger, fear, sadness, boredom, and anxiety.

Studies have shown that people speak happiness (joy) and fear at higher frequency ranges (heard as pitch) than emotions such as sadness. Anger seems to be an emotion that can go in two directions, phonetically:

- **Hot anger:** When people go up high with the voice and show much variability.
- **Cold anger:** When people are brooding with low pitch range, high intensity, and fast *attack times* (sudden rise in amplitude) at voice onset.

Emotional patterns in speech (also known as *affective prosody*) don't directly affect sentence meaning. However, these patterns can interact with linguistic prosody to affect listeners' understanding. For instance, adults with cerebral right hemisphere damage (RHD) can have difficulty understanding, producing, and mimicking the emotional components of speech. The speech of such individuals can often be *monotonic* (flat). It can sometimes be challenging for clinicians to sort out which aspects of these speech presentations are due to emotion or to linguistic deficits.

Fine-Tuning Speech Melodies

Phoneticians can be sticklers for detail. They just don't like messy bits left over. In addition to the different types of stress, intonation, focus, and emotional prosody, certain aspects of speech melody still require measures to account for them. These sections examine two such measures.

Sonority: A general measure of sound

Sono- means sounds, and *sonority* is therefore a measure of the relative amount of sound something has. Technically, *sonority* refers to a sound's loudness relative to those of other sounds having the same length, stress, and pitch. This measure of sound is particularly handy for working with tone languages, such as Vietnamese, where decisions about tone structure are important.

To get a clearer sense of this jargon, try saying the sound "a" (/ɑ /) followed by the sound "t." Assuming you spoke them at the same rate and loudness, the vowel /ɑ/ should be much more sonorous (have more sound) than the voiceless stop, "t."

The concept of sonority is relative, which means phoneticians often refer to sonority hierarchies or scales. In a *sonority hierarchy*, classes of sounds are grouped by their degree of relative loudness. Check out `www-01.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsTheSonorityScale.htm` for an example of one.

A sonority scale expresses more fine-grained details. For instance, according to phonologist Elizabeth Selkirk, English sounds show the following ranking:

([ɑ] > [e=o] > [i=u] > [r] > [l] > [m=n] > [z=v=ð] > [s=f=θ] > [b=d=g] > [p=t=k])

If you try out some points on this scale, you'll hear, for example, that [ɑ] is more sonorous than [i] and [u].

Sonority is an important principle regulating many phonological processes in language, including *phonotactics* (permissible combinations of phonemes) syllable structure, and stress assignment.

Prominence: Sticking out in unexpected ways

When all is said and done, some problem cases of prosody can still challenge phoneticians. One such problem is exactly how stress is assigned to syllables in words. For instance, some English words can be produced with different amounts of syllables. Consider the words "frightening" and "maddening."

Do you say them with two syllables, such as /ˈfɹaɪtnɪŋ/ and /ˈmædnɪŋ/? Or do you use three syllables, such as /ˈfɹaɪtənɪŋ/ and /ˈmædənɪŋ/? Or sometimes with two and sometimes with three?

Other English words may change meaning based on whether they are pronounced with two or three syllables. For instance:

"lightning" (such as in a storm) /ˈlaɪtnɪŋ/

"lightening" (such as, getting brighter) /ˈlaɪtənɪŋ/

A proposed solution for the more difficult cases of stress patterns is to rely on a feature called *prominence*, consisting of a combination of sonority, length, stress, and pitch. According to this view, prominence peaks are heard in words to define syllables, not solely sonority values.

Prominence remains a rather complex and controversial notion. It's an important concept in *metrical phonology* (a theory concerned with organizing segments into groups of relative prominence), where it's often supported with data from speech experiments. However, other phoneticians have suggested different approaches may be more beneficial in addressing the problems of syllabicity in English (such as the application of speech technology algorithms, rather than linguistic descriptions).

Chapter 11

Marking Melody in Your Transcription

In This Chapter

- Sampling choices for prosodic transcribing
- Defining the tonic syllable and intonational phrase
- Becoming proficient at a three-step process
- Rising and tagging

Imagine you're sitting in a busy restaurant in a big city hearing many different foreign languages spoken. It's noisy, but you want to impress your friends with your (amazing) ability to tell which language is which. One important clue to help you is *language melody,* which includes *stress* (when a syllable is louder, longer, and higher because the talker uses extra breath) and *intonation* (a changing tune during a phrase or sentence).

For instance, someone speaking Spanish has a very different melody than someone speaking Mandarin, and you can hear it if you know what you're listening for. However, capturing these details in written transcription is much more difficult, particularly if you need to compare healthy and disordered speech.

In this chapter, I show you some practical ways to incorporate melodic detail in your transcriptions. I begin with a tried-and-true method useful for clinical notes or field transcription. I also include some examples of a more systemized method, the Tone and Break Indices (ToBI) that linguists and many people in the speech science community use.

Focusing on Stress

When transcribing many languages, being able to identify a stressed syllable is essential. Knowing these characteristics of a stressed syllable can help

you identify it. An English *stressed syllable* is louder, longer (in duration), and higher in pitch. In English, stress plays a number of important roles:

- At the vocabulary level, *polysyllabic* words (with more than one syllable) have specified stress that a native speaker must correctly produce to sound appropriate. Thus, "*syl*lable" is okay, but "syl*lab*le" sounds weird.
- For word function, stress makes a difference between nouns such as "*re*bel" and verbs, such as "to re*bel*."
- In phrases and sentences, stress changes *focus*, or emphasis. For example, although these two sentences contain the exact same words, stressing different words gives a different emphasis:

  "*She* never wears Spandex!" (He does!)

  "She never *wears* Spandex!" (She sells it, instead.)

Stress also plays a special role in English when it serves as the *tonic syllable* (a syllable that stands out because it carries the major pitch change of a phrase or sentence).

The following sections describe some of the complexities involved in speech that can make the job of transcribing language melody a challenge.

Recognizing factors that make connected speech hard to transcribe

Understanding the role that stress plays in English is important, and a futher challenge is to be able to accurately complete a prosodic transcription of connected speech.

Transcribing *prosody* (the melody of language) can be challenging, for a number of reasons:

- Several types of prosodic information are present in a person's speech. This information includes *linguistic prosody* where melody and timing specifically affect language, as well as *emotional prosody*, reflecting the speaker's mood and attitude toward what the speaker is discussing.
- People don't usually speak in complete sentences. Nor do they always cleanly break at word or phrase boundaries. For example, here is some everyday talk from teenagers in Dallas, Texas: "So, like, I was gonna see this movie at North Park? But then Alex was there? So . . . yeah, and . . . then it's like . . . Awkward!" (This is an example of the Valley Girl social dialect; refer to Chapter 18 for more information.)

Listen to yourself talk sometime, and tune in to the grammatical structures that you use and the precision with which you articulate. Speaking in different *registers* (level of language used for a particular setting) in different settings is natural. When presenting professionally (such as in class, work, or clinic), people are usually on their best behavior and tend to use complete sentences, full grammatical constructions, and more fully achieved articulatory targets (called *hyperspeech*). In contrast, when talking casually with friends, people naturally relax and use more informal constructions, centralized vowels, and reduced articulatory precision, referred to as *hypospeech*.

Finding intonational phrases

The IPA doesn't recommend any one system for capturing language melody (prosody). Instead, various phoneticians have applied rules and theories in the best ways they see fit. Fortunately, many methods are available. One time-honored method begins with defining an intonational phrase. Based on these building blocks, you, as a transcriber, can achieve different degrees of success.

An *intonational phrase,* sometimes called a *tone unit, tonic phrase,* or *tone group,* is a pattern of pitch changes that matches up in a meaningful way with a part of a sentence. Although the exact definition varies between phoneticians, certain key characteristics of an intonational phrase are as follows:

- A part of connected speech containing one tonic syllable.
- Similar to a *breath-group* (sequence of sounds spoken in a single exhalation), a single, continuous airstream supports it.
- Similar to a phrase, a clause, or a non-complex sentence.
- Similar to breaks signaled by written punctuation (commas, periods, or dashes.)
- Intonational phrases aren't syntactic units, but they can frequently match up to them in a practical sense.

Check out these examples:

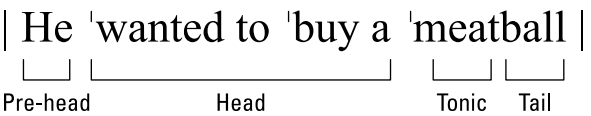
Example Words	*Number of Intonational Phrases*
|ˈYep!|	1
|The ˈdog.|	1
|Although he ignored the ˈcat,|the boy fed the ˈdog.|	2
|The boy fed the ˈdog,|but ignored the ˈcat.|	2
|The boy fed the ˈdog, | gave it a ˈmeatball,|but ignored the ˈcat.|	3

In these examples, the boundaries of intonational phrases are divided using vertical lines ([|]). The words that typically receive stress have a primary stress mark ([ˈ]) before them. Single words (such as "Yep") and fragments ("The dog") can be intonational phrases. If the speaker is communicating too much in a single breath-group, the utterance is often broken down into separate, shorter tone units (such as phrases, clauses, or shorter bits of choppier speech) containing between one to three intonational phrases (shown here). It's common for a spoken sentence to have one to two intonational phrases, but there could be more, depending on how a person is talking.

Zeroing in on the tonic syllable

Each intonational phrase will have one (and only one) tonic syllable (also called the *nuclear syllable*), the syllable that carries the most pitch change. The tonic syllable is an important idea for many theories of prosody.

A *tonic syllable* is the key part of an intonational phrase because it's a starting point for the melody of that phrase. Together, the concepts of a tonic syllable and intonational phrase allow a thorough description of language melody. If this theory sounds circular to you, it is. However, it's by intention. Here is how it works: An intonational phrase consists of a nucleus and an optional pre-head, head, and tail. The following figure shows an example for an intonational phrase.



Taken separately or in combination these components can describe English melody. This tonic syllable/intonational phrase system is often used for teaching students of English as a second language. It's particularly well suited for British English, especially the Received Pronunciation (RP) accent.

Seeing how phoneticians have reached these conclusions

Phoneticians have come up with these explanations of English melody by considering several factors, including the rhythm of English (called *meter*, described in units called *feet*). Phoneticians also note that intonation corresponds with different types of meaning, such as statements and questions (refer to the next section for more information).

It's beyond the scope of introductory phonetics to explain these theories of prosody. However, you should be able to form an intuitive sense of what an intonational phrase is. For instance, if you review the examples in the previous section, "Finding intonational phrases," you can hear that "The dog" receives a lot of stress, whereas other parts of the sentences (such as "although") don't receive much stress.

Consider the sentence "The boy fed the dog." You can pronounce this sentence in many different ways, depending on what you're emphasizing (for example, "The *boy* fed the dog," "The boy *fed* the dog," and so on). In these cases (which some phoneticians call a *dislocated tonic*), emphasis or focus has shifted the position of the tonic syllable. However, in most cases, the tonic syllable of an intonational phrase is the last stressed syllable that conveys new information, such as:

"The boy fed the *dog*" /ðə bɔɪ fɛd ðə ˈdɔg/

Here, "dog" is the tonic syllable, carrying the most prosodic information.

Sometimes a person doesn't produce intonational phrases in the usual manner. In actual transcriptions, you may encounter speech like this:

"The boy . . . fed . . . the dog." /ðə ˈbɔɪ|ˈfɛd | ðə ˈdɔg/

This type of speech (*hesitant speech*) would have more intonational phrases. This particular example uses three intonational phrases instead of one. The tonic syllables would be "boy," "fed," and "dog."

Consider cases where stress is changed due to emphasis. If someone is excited about the fact that the dog was *fed*, rather than *washed*, he or she would probably say the following:

"The boy *fed* the dog" /ðə bɔɪ ˈfɛd ðə dɔg/

This time, "fed" is the tonic syllable of a single intonational phrase.

Applying Intonational Phrase Analysis to Your Transcriptions

Being able to apply intonational phrase analysis can give you a better idea of how phoneticians handle the challenge of transcribing intonation and prosody. This method demonstrates an accurate and easy-to-complete method of prosodic transcription. Although this method has its limitations of describing prosody because it doesn't provide fine-grained details such as the subcomponents of a tonic phrase (pre-head, head, nucleus, and tail), together with

narrow transcription (recording details about phonetic variations and allophones), it does provide an easy way of denoting the melody of connected speech.

Here I walk you through these three steps and use this example to explain this process:

> "The earliest phoneticians were the Indian grammarians."
>
> /ðɪ ɝlɪɛst fɔnətɪʃənz wɚðə ɪndɪən gɹəmɛɹɪənz/
> (The broad transcription, with no details yet filled in.)
>
> If you want to listen to the sound file, check it out at `www.utdallas.edu/~wkatz/PFD/the_earliest_phonetician_WK.wav`. This is a recording of me reading a passage in a matter-of-fact manner.

1. **Locate prosodic breaks corresponding with the breath groups.**

   To find them, listen for clear gaps during speech. After you locate them, place a vertical bar ([|]) for minor phrase breaks and a double-bar ([‖]) for major phrase-breaks.

   For this example, your work should look like this:

   /ðɪ ɝlɪɛst fɔnətɪʃənz | wɚðə ɪndɪən gɹəmɛɹɪənz‖/

2. **Mark the tonic syllable in each tone unit (intonational phrase) as the primary stressed syllable and denote the stress in other polysyllabic words by marking them with secondary stress.**

   Mark the stress of "ˈ*ear*liest" and "graˈ*mmar*ian" with a primary stress mark. In this case, the stress mark further indicates the tonic syllable of an intonational phrase.

   At this stage, your transcription should look like this:

   /ðɪ ˈɝlɪɛst fɔnətɪʃənz |wɚðə ɪndɪən gɹəˈmɛɹɪənz‖/

   Continue to mark stress on the other polysyllabic words ("phone*ti*cians" and "*In*dian") to produce the following:

   /ðɪ ˈɝlɪɛst fɔnəˌtɪʃənz |wɚðə ˌɪndɪən gɹəˈmɛɹɪənz‖/

3. **Draw an estimate of the fundamental frequency contour *(pitch plot)* above the IPA transcription.**

   This is the best part. Use your ear (and hand) to draw the shape of the intonation contour above the transcription. This task is rather like *follow the bouncing ball.* Refer to Figure 11-1 for an example.

   In this figure, a hand-drawn pitch contour marks sounds going up and sounds going down. These plots are helpful for transcribing the intonational phrases of connected speech.

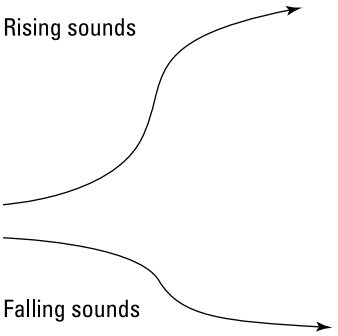


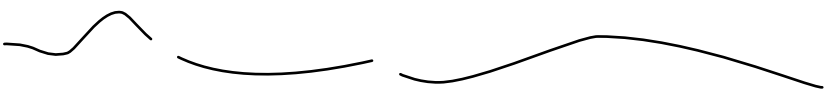
Illustration by Wiley, Composition Services Graphics

Figure 11-1: Rising sounds go up, falling sounds go down.

If you're musically or artistically challenged, you can build your confidence for intonation contour sketching in these ways:

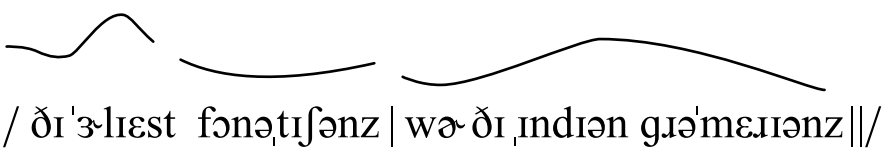
- **Practice.** The old saying is true: Practice makes perfect.
- **Use a speech analysis program.** These types of programs, such as WaveSurfer or Praat (Dutch for "Speech"), can help you analyze the fundamental frequency patterns of the utterances you want to transcribe and compare your freehand attempts with the instrumental results. You're probably better than you think you are.

The intonation contour you arrive at should look something like this:



Don't worry whether you've smoothly connected the pitch contour or whether you make less connected straight forms. The main point is that your figure rises when the pitch rises (such as during the word "earliest") and falls appropriately (as in the end of the phrase). The goal of using this three-step method is to uncover the melody of the original utterance.

When you finish, your transcription should look like this:



For more practice, go in the opposite direction. Look up `www.utdallas.edu/~wkatz/PFD/the_earliest_phonetician_WK2.wav` for a link to a sound file of the phrase, "The earliest phoneticians were the Indian grammarians" read by yours truly in a slightly different manner. This time, I use an impatient tone of voice. Your job is to produce a narrow transcription marked with intonational phrases, tonic syllables, and a prosodic contour (follow the three-step method to do so). Good luck!

You can check your work by going to `www.utdallas.edu/~wkatz/PFD/the_earliest_phoneticians_answer.gif`.

Tracing Contours: Continuation Rises and Tag Questions

Chapter 10 discusses the three main patterns for English sentence-level intonation. Two other common intonation patterns exist. They can differ slightly as a function of dialect (American versus British English) as well as the mood and attitude of the speaker. These sections take a closer look specifically at continuation rises and tag questions to show you where they occur and how to transcribe them.

Continuing phrases with a rise

A *continuation rise* is a conspicuous lack of a falling pattern on the tonic syllable at the end of that phrase. It occurs when one intonational phrase follows another. For instance, contrast the falling pattern on the tonic syllable ("crazy") in the first example sentence and the continuation rise for that same word in the second example.

- "Eileen is really *crazy*." (`www.utdallas.edu/~wkatz/PFD/eileen_crazy1.wav`)
- "Eileen is really crazy, but she's my best friend." (`www.utdallas.edu/~wkatz/PFD/eileen_crazy2.wav`)

Continuation rise patterns are common in English lists. Here is a (ridiculously healthy) shopping list: "She bought peaches, apples, and kiwis." Most North American English speakers pronounce it something like what appears in Figure 11-2.

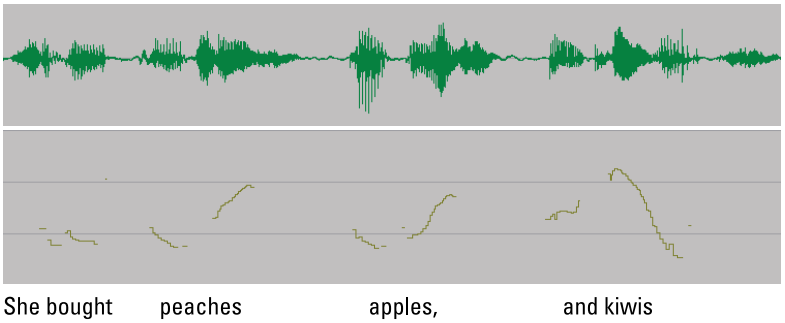


Figure 11-2: The speech waveform (above) and the intonation contour (below) of "peaches, apples, and kiwis."

Illustration by Wiley, Composition Services Graphics

In this figure, notice that the words "peaches" and "apples" rise during the continuation of the sentence, but the word "kiwis" falls at the end. If you were to flip this order and use falling prosody during the production (for "peaches" and "apples"), while rising at the end (for "kiwis"), you would sound, frankly, bizarre.

You have a talker who coughs and says "um" and "er" during the transcription. What do you do? You may wonder if you have to put those utterances in your transcription. For most narrow transcriptions, the answer is yes. Filled pauses such as [əm] and [ɚ] are common in speech. Talkers with foreign accents may use very different filled pauses than native speakers of English (for instance, [ɛm] for Hebrew speakers and [eː] or [n̩] for Japanese speakers). You can indicate nonlinguistic vocalizations (such as coughing, sneezing, laughter) in parentheses.

Tagging along

English *tag questions* (statements made into a question by adding a fragment at the end) have their own characteristic patterns. Tag questions can be either rising or falling. Their patterns depend somewhat on the dialect used (for instance, British or American), but mainly they depend on the exact use of the tag.

Rising patterns are found when a tag question turns a statement into a question, such as these examples:

- "You're kidding, *aren't you?*"
- "It's a real Rolex, *isn't it?*"

Falling patterns are used to emphasize a statement that was just made:

- "He sold you a fake Rolex, *didn't he?*"
- "That's really awful, *isn't it?*"

Testing Out ToBI

The *tone and break indices system (ToBI)* is a set of conventions used for working with speech prosody. ToBI is mainly designed for English, although conventions are being developed for other languages (including German, Korean, Japanese, and Greek). Researchers primarily use this system, and it's a bit more advanced for most clinical and educational settings. However, having a basic grasp of ToBI is important because you'll probably encounter some literature referring to it.

Here are some characteristics concerning ToBI that are important to grasp:

- Instead of drawing a pitch contour over a transcription, ToBI describes pitch peaks and valleys in terms of high and low *target tones,* which are represented as combinations of the letters H (for high) and L (for low). A target tone can be simple, such as *high* (written as H*, called *H star*), or it can have gliding properties, such as L+H*. In this case, the tone starts low (L), and then glides up to high (H*).

  For example, if someone asked "Well?", the target tone would be L+H*.

- Target tones are typically written on a line of text (called a *tier*). A second tier represents break indices, showing pause and gap durations. ToBI uses a range of boundary strength levels, from 0 to 4, representing shortest to longest. A break of 0 represents no break (such as in the contraction between *we* and *are* — w*e're*), a 1 represents most breaks between words, and breaks of 3 and 4 indicate intentional breaks in a phrase and at a sentence ending.

- In ToBI, the last pitch accent of a phrase is called the *nuclear pitch accent*, which is similar to the idea of a tonic syllable. For instance, in a straightforward reading of "The boy fed the dog," the nuclear pitch accent is the word "dog."

- ToBI also allows each phrase to receive another marking after nuclear pitch, called a *phrase accent*. These markers (L- and H-) permit additional prosodic refinement.

- A boundary tone (L% or H%) acts as a kind of marker for sentence-level intonation. When the sentence has an overall falling intonation, as in a simple declarative pattern, the boundary tone is L%. When pitch rises, as in a yes-no question or continuation rise, the sentence has an H% boundary tone. Boundary tones are placed at phrase edges.

Check out this example that compares tonic syllable analysis (that I discuss in this chapter) with ToBI for the sentence, "The boy fed the *dog.*"

- **Tonic syllable analysis:**
  /|ðə bɔɪ fɛd ðə ˈdɔg||/

  Here, the word "dog" is the tonic syllable of an intonational phrase.

- **ToBI analysis:**

Break index	[	1	1	1	4	]
Tone tier	[	H*		H*L-L%		
Segmental tier	[ðə	bɔɪ	fɛd	ðə	ˈdɔg	]

The ToBI analysis specifies level 1 breaks (those used for most middle-of-phrase boundaries) between the words of the sentence and a phrase level break (4) at the end. The tone tier shows high nuclear tones throughout, followed

by a nuclear accent marked with a low phrase accent (L-) and a low boundary tone (L%), indicating phrase final fall.

The other H* marker ("boy") is an optional prenuclear pitch accent. In this case, it indicates that the entire utterance is produced with a high, flat intonation (with a final fall). However, more gradual pitch declination, called *downdrift*, can be indicated by adding *downstepping* symbols, !H*.

For instance, a ToBI representation of a more declining pattern could be represented like this:

Break index	[1 1 1 1 4]
Tone tier	[H* !H* !H*L-L%]
Segmental tier	[ðə bɔɪ fɛd ðə ˈdɔg]

If you're using a program such as Praat or WaveSurfer, you can integrate ToBi, labeling directly into your graphics. Refer to `www.cs.columbia.edu/~agus/tobi/` for more information on incorporating ToBI into waveform editing packages.

Part III

Having a Blast: Sound, Waveforms, and Speech Movement

"My husband's uvula is shaped like a duck call. When he sneezes, a flock of mallards usually shows up."

web extras

Visit dummies.com/extras/phonetics for more great Dummies content online.

In this part . . .

- Comprehend what causes sound and know why this information is essential for understanding how people talk and listen.
- Grasp how to describe sound physically, in terms of frequency, amplitude, and duration.
- Be able to relate physical aspects of sound to people's subjective listening patterns.
- Know how to decode the information in sound spectrograms.
- Gather the basics of current models of human speech perception.

Chapter 12

Making Waves: An Overview of Sound

In This Chapter

- Working with sound waves
- Getting grounded in the physics needed to understand speech
- Relating sound production to your speech articulators

One of the great things about phonetics is that it's a bridge to fields like acoustics, music, and physics. To understand speech sounds, you must explore the world of sound itself, including waves, vibration, and resonance. Many phoneticians seem to be musicians (either at the professional level or as spirited amateurs), and it's normal to find phoneticians hanging around meetings of the Acoustical Society of America. Just trying to talk about this accent or that isn't good enough; if you want to practice good phonetics, you need to know something about acoustics.

This chapter introduces you to the world of sound and describes some basic math and physics needed to better understand speech. It also explains some essential concepts useful for analyzing speech with a computer.

Defining Sound

Sound refers to energy that travels through the air or another medium and can be heard when it reaches the ear. Physically, sound is a longitudinal wave (also known as a *compression wave*). Such a wave is caused when something displaces matter (like somebody's voice yelling, "Look out for that ice cream truck!") and that vibration moves back and forth through the air, causing compression and *rarefaction* (a loss of density, the opposite of compression). When this pressure pattern reaches the ear of the listener, the person will hear it.

When a person shouts, the longitudinal wave hitting another person's ear demonstrates compression and rarefaction. The air particles themselves don't actually move relative to their starting point. They're simply the medium that the sound moves through. None of the air expelled from the person shouting about the truck actually reaches the ear, just the energy itself. It's similar to throwing a rock in the middle of a pond. Waves from the impact will eventually hit the shore, but this isn't the water from the center of the pond, just the energy from the rock's impact.

The speed of sound isn't constant; it varies depending on the stuff it travels through. In air, depending on the purity, temperature, and so forth, sound travels at approximately 740 to 741.5 miles per hour. Sound travels faster through water than through air because water is denser than air (the denser the medium, the faster sound can travel through it). The problem is, humans aren't built to interpret this faster signal in their two ears, and they can't properly pinpoint the signal. For this reason, scuba diving instructors train student divers not to trust their sense of sound localization underwater (for sources such as the dive boat motor). It is just too risky. You can shout at someone underwater and be heard, although the person may not be able to tell where you are.

Cruising with Waves

The universe couldn't exist without waves. Most people have a basic idea of waves, perhaps from watching the ocean or other bodies of water. However, to better understand speech sounds, allow me to further define waves and their properties.

Here are some basic facts about sound and waves:

- Sound is energy transmitted in longitudinal waves.
- Because it needs a medium, sound can't travel in a vacuum.
- Sounds waves travel through media (such as air and water) at different speeds.
- *Sine* (also known as *sinusoid*) waves are simple waves having a single peak and trough structure and a single (fundamental) frequency. The fundamental frequency is the basic vibrating frequency of an entire object, not of its fluttering at higher harmonics.
- People speak in complex waves, not sine waves.

- Complex waves can be considered a series of many sine waves added together.
- Fourier analysis breaks down complex waves into sine waves (refer to the sidebar later in this chapter for more information).
- Complex waves can be *periodic* (as in voiced sounds) or *aperiodic* (as in noisy sounds). Check out the "Sine waves" and "Complex waves" sections for more on periodic and aperiodic waves.

These sections give examples of simple and complex waves, including the relation between the two types of waveforms. I also describe some real-world applications.

Sine waves

The first wave to remember is the *sine wave* (or sinusoid), also called a *simple wave*. Sine is a trigonomic function relating the opposite side of a right-angled triangle to the hypotenuse.

There are some good ways to remember sine waves. Here is a handy list:

- Sine waves are the basic building blocks of the wave world.
- All waveforms can be broken down into a series of sine waves.
- Many things in nature create sine waves — basically anything that sets up a simple oscillation. Figure 12-1 shows a sine wave being created as a piece of paper is pulled under a pendulum that's swinging back and forth.
- In western Texas, if you're lucky, you may see a beautiful sine wave in the sand left by a sidewinder rattlesnake.
- When sound waves are sine waves, they're called *pure tones* and sound cool or cold, like a tuning fork or a flute (not a human voice or a trumpet). This is because the physics of sine wave production involve emphasizing one frequency, either by forcing sound through a hole (as in a flute or whistle) or by generating sound with very precisely machined arms (which reinforce each other as they vibrate), in the case of the tuning fork).

Sine waves are used in clinical audiology for an important test known as *pure-tone audiometry*. Yes, those spooky tones you sometimes can barely hear during an audiology exam are sine waves designed to probe your threshold of hearing. This allows the clinician to rule out different types of hearing loss.

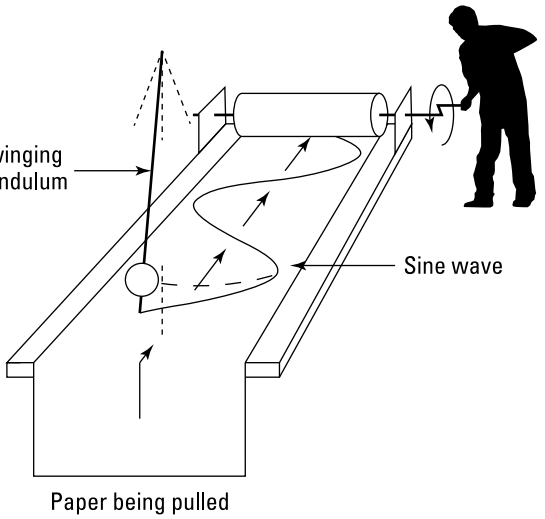


Figure 12-1: A pendulum creating sine waves on a piece of paper being pulled by an enthusiastic phonetician.

Illustration by Wiley, Composition Services Graphics

Complex waves

Everyone knows the world can be pretty complex. Waves are no exception. Unless you're whistling, you don't produce simple waves — all your speech, yelling, humming, whispering, or singing otherwise consists of complex wave production.

A *complex wave* is like a combination of sine waves all piled together. To put it another way, complex waves have more than one simple component — they reflect several frequencies made not by a simple, single vibrating movement (one pendulum motion) but by a number of interrelated motions. It's similar to the way that white light is complex because it's actually a mixture of frequencies of pure light representing the individual colors of the rainbow.

Getting into the formula of sine

If you like formulas, sine waves are created by the sine function:

$$y(t) = A\sin(2\pi ft + \phi)$$

In this formula:

- ***A:*** The *amplitude* is the peak deviation of the function from zero.
- ***f:*** The *frequency* is the number of oscillations (cycles) that occur each second of time.
- **ϕ:** The *phase* specifies (in radians) where in its cycle the oscillation is at $t = 0$.

If you aren't a math fan, no worries!

Measuring Waves

Every wave can be described in terms of its *frequency, amplitude,* and *duration*. But when two or more waves combine, *phase* comes into play. In this section, you discover each of these terms and what they mean to sound.

Frequency

Frequency is the number of times something happens, divided by time. For instance, if you go to the dentist twice a year, your frequency of dental visits is two times per year. But sound waves repeat faster and therefore have a higher frequency.

Frequency is a very important measure in acoustic phonetics. The number of cycles per second is called *hertz* (Hz) after the famous German physicist Heinrich Hertz. Another commonly used metric is kilohertz (abbreviated kHz), meaning 1,000 Hertz. Thus, 2 kHz = 2,000 Hz = 2,000 cycles per second.

The range of human hearing is roughly 20 to 20,000 cycles per second, which means that the rate of repetition for something to cause such sound is 20 to 20,000 occurrences per second. A bullfrog croaks in the low range (fundamental frequency of approximately 100 Hz), and songbirds sing in the high range (the house sparrow ranges from 675 to 18,000 Hz).

Figure 12-2 shows a sample of frequency demonstrated with a simple example so that you can count the number of oscillations and compute the frequency for yourself. In Figure 12-2a (periodic wave), you can see that the *waveform* (the curve showing the shape of the wave over time) repeats once in one second (shown on the *x*-axis). Therefore, the frequency is one cycle per second, or 1 Hz. If this were sound, you couldn't hear it because it's under the 20 to 20,000 Hz range that people normally hear.

Seeing is believing! An oscillation can be counted from peak to peak, valley to valley, or zero-crossing to zero-crossing.

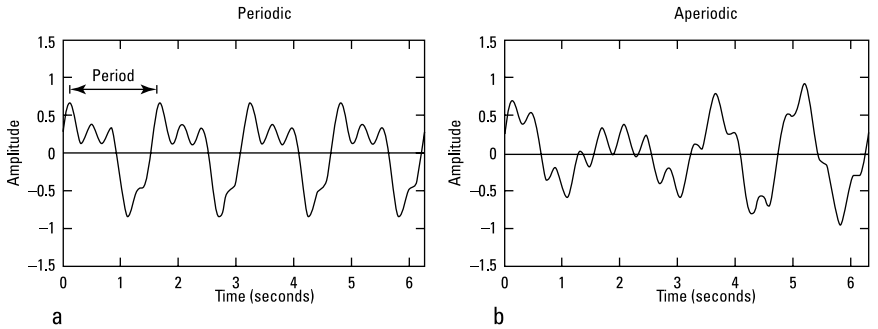


Figure 12-2: A sample periodic wave (a) and an aperiodic wave (b).

Illustration by Wiley, Composition Services Graphics

Period is a useful term related to frequency — it's a measure of the time between two oscillations and the inverse of frequency. If your frequency of dental visits is two times per year, your period of dental visits is every six months.

Waves produced by irregular vibration are said to be *periodic*. These waves sound musical. Sine waves are periodic, and most musical instruments create periodic complex waves. However, waves with cycles of different lengths are *aperiodic* — these sound more like noise. An example would be clapping your hands or hearing a hissing radiator. Figure 12-2b shows an aperiodic wave.

You can also talk about the length of the wave itself. You can sometimes read about the wavelength of light, for example. But did you ever hear about the wavelength of sound? Probably not. This is because wavelengths for sound audible to humans are relatively long, from 17 millimeters to 17 meters, and are therefore rather cumbersome to work with. On the other hand, sound wavelength measurements can be handy for scientists handling higher frequencies, such as ultrasound, which uses much higher frequencies (and therefore much shorter wavelengths).

One frequency that will come in very handy is the *fundamental frequency*, which is the basic frequency of a vibrating body. It's abbreviated F_0 and is often called *F-zero* or *F-nought*. A sound's fundamental frequency is the main information telling your ear how low or high a sound is. That is, F_0 gives you information about *pitch* (see the section "Relating the physical to the psychological" in this chapter).

Amplitude

Amplitude refers to how forceful a wave is. If there is a weak, wimpy oscillation, there will be a tiny change in the wave's amplitude, reflected on the vertical axis. Such a wave will generally sound quiet. Figure 12-3 shows two waves with the same frequency, where one (shown in the solid line) has twice the amplitude as the other (shown in the dotted line).

Sound amplitude is typically expressed in terms of the air pressure of the wave. The greater the energy behind your yell, the more air pressure and the higher the amplitude of the speech sound. Sound amplitude is also frequently described in decibels (dB). Decibel scales are important and used in many fields including electronics and optics, so it's worth taking a moment to introduce them.

In the following list, I give you the most important things about dB you need to know:

- One dB = one-tenth of a bel.

  The bel was named after Alexander Graham Bell, father of the telephone, which was originally intended as a talking device for the deaf.

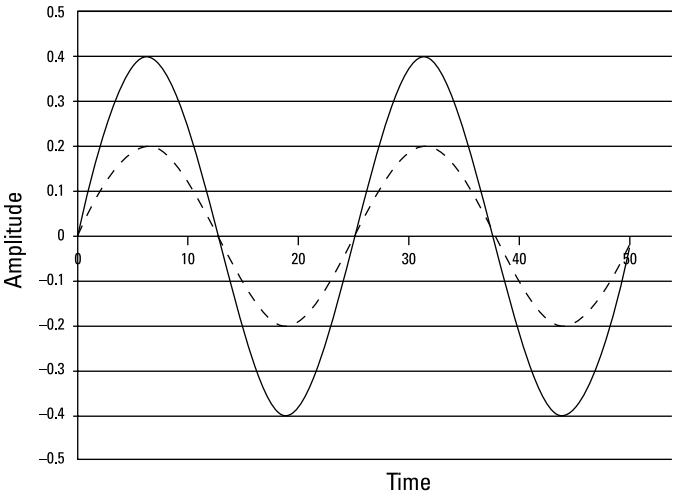


Figure 12-3: Two waveforms with the same frequency and different amplitude.

Illustration by Wiley, Composition Services Graphics

- dB is a logarithmic scale, so an increase of 10 dB represents a ten-fold increase in sound level and causes a doubling of perceived loudness.

  In other words, if the sound of one lawnmower measures 80 dB, then 90 dB would be the equivalent sound of ten lawnmowers. You would hear them twice as loud as one lawnmower.

- Sound levels are often adjusted (weighted) to match the hearing abilities of a given critter. Sound levels adjusted for human hearing are expressed as dB(A) (read as "dee bee A").

  The dBA scale is based on a predefined threshold of hearing reference value for a sine wave at 1000 Hz — the point at which people can barely hear.

- Conversational speech is typically held at about 60 dBA.

- Too much amplitude can hurt the ears. Noise-induced hearing damage can result from sustained exposure to loud sounds (85 dB and up).

A property associated with amplitude is *damping,* the gradual loss of energy in a waveform. Most vibrating systems don't last forever; they peter out. This shows up in the waveform with gradually reduced amplitude, as shown in Figure 12-4.

Duration

Duration is a measure of how long or short a sound lasts. For speech, duration is usually measured in seconds (for longer units such as words, phrases, and sentences) and *milliseconds* (ms) for individual vowels and consonants.

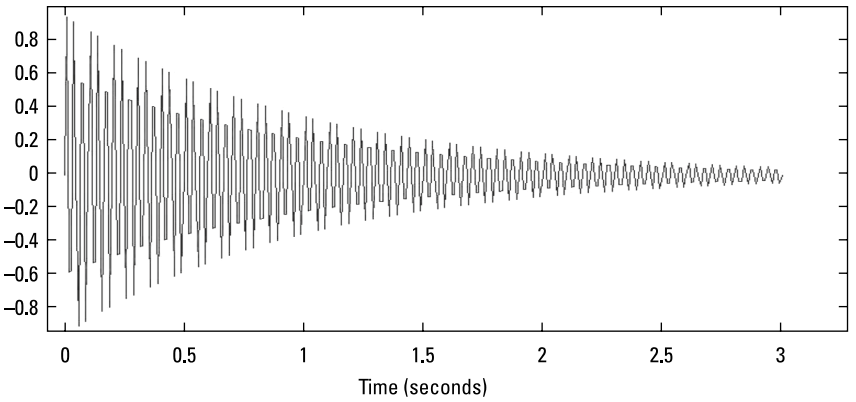


Figure 12-4: Damping happens when there is a loss of vibration due to friction.

Illustration by Wiley, Composition Services Graphics

Phase

Phase is a measure of the time (or angle) between two similar events that run at roughly the same time. Phase can't be measured with a single sound — you need two (waves) to tango. Take a look at Figure 12-5 to get the idea of how it works:

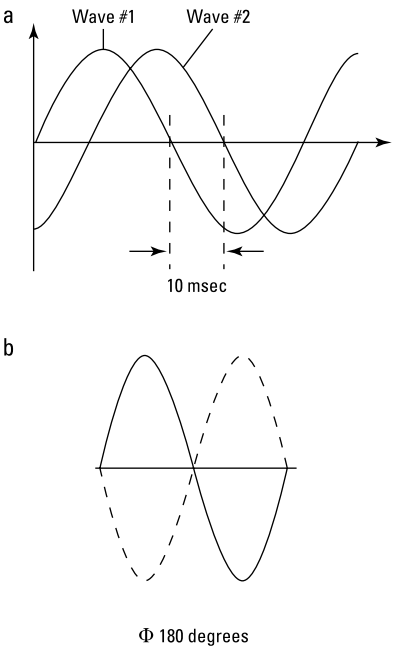


Figure 12-5: Two examples of phase differences — by time (a) and by angle (b).

Illustration by Wiley, Composition Services Graphics

In the top example of Figure 12-5, when wave #1 starts out, wave #2 lags by approximately 10 msec. That is, wave #2 follows the same pattern but is 10 msec behind. This is phase described by time.

The bottom example in Figure 12-5 shows phase described by angle. Two waves are 180 degrees out of phase. This example is described by *phase angle,* thinking of a circle, where the whole is 360 degrees and the half is 180 degrees. To be *180 degrees out of phase* means that when one wave is at its peak, the other is at its valley. It's kind of like a horse race. If one horse is a quarter of a track behind the other horse, you could describe him as being so many yards, or 90 degrees, or a quarter-track behind.

Relating the physical to the psychological

In a perfect world, what you see is what you get. The interesting thing about being an (imperfect) human is that the physical world doesn't relate in a one-to-one fashion with the way people perceive it. That is, just because something vibrates with such and such more energy doesn't mean you necessarily hear it as that much louder. Settings in your perceptual system make certain sounds seem louder than others and can even set up auditory illusions (similar to optical illusions in vision).

This makes sense if you consider how animals are tuned to their environment. Dogs hear high-pitched sounds, elephants are tuned to low frequencies (infrasound) for long-distance communication, and different creatures have different perceptual settings in which trade-offs between frequency, amplitude, and duration play a role in perception. Scientists are so intrigued by this kind of thing that they have made it into its own field of study — *psychophysics,* which is the relationship between physical stimuli and the sensations and perceptions they cause.

Pitch

The psychological impression of fundamental frequency is called *pitch*. High-frequency vibrations sound like high notes, and low-frequency vibrations sound like low notes. The ordinary person can hear between 20 to 20,000 Hz. About 30 to 35 percent of people between 65 and 75 years of age may lose some hearing of higher-pitched sounds, a condition called *presbycusis* (literally "aged hearing").

Loudness

People hear amplitude as *loudness,* a subjective measure that ranges from quiet to loud. Although many measures of sound strength may attempt to adjust to human loudness values, to really measure loudness values is a complex process — it requires human listeners.

Different sounds with the very same amplitude won't have the same loudness, depending on the frequency. If two sounds have the same amplitude and their frequencies lie between about 600 and 2,000 Hz, they'll be perceived to be about the same loudness. Otherwise, things get weird! For sounds near 3,000 to 4,000 Hz, the ear is extra-sensitive; these sounds are perceived as being louder than a 1,000 Hz sound of the same amplitude. At frequencies lower than 300 Hz, the ear becomes less sensitive; sounds here are perceived as being less loud than they (logically) "should" be.

This means I can freak you out with the following test. I can play you a 300 Hz tone, a 1,000 Hz tone, and a 4,000 Hz tone, all at *exactly* the same amplitude. I can even show you on a sound-level meter that they are exactly the same. However, although you know they are all the same, you'll *hear* the three as loud, louder, and loudest. Welcome to psychophysics.

Length

The psychological take on duration is *length*. The greater the duration of a speech sound, the longer that signal generally sounds. Again, however, it's not quite as simple as it may seem. Some languages have sounds that listeners hear as double or twin consonants. (***Note:*** Although English spelling has double "n," "t," and so forth, it doesn't always pronounce these sounds for twice as long.) Doubled consonant sounds are called *geminates* (twins). It turns out that geminates are usually about twice the duration as nongeminates. However, it depends on the language. In Japanese, for example, geminates are produced about two to three times as long as nongeminates. An example is /hato/ "dove" versus /hatto/ "hat."

Sound localization

Humans and other creatures use phase for *sound localization,* which allows them to tell where a sound is coming from. A great way to test whether you can do this is to sit in a chair, shut your eyes, and have a friend stand about 3 feet behind you. Have her snap her fingers randomly around the back and sides of your head. Your job is to point to the snap, based only on sound, each time.

Most people do really well at this exercise. Your auditory system uses several types of information for this kind of task, including the time-level difference between the snap waveform hitting your left and right ears, that is — phase. After more than a century of work on this issue, researchers still have a lot to learn about how humans localize sound. There are many important practical applications for this question, including the need to produce better hearing aids and communication systems (military and commercial) that preserve localization information in noisy environments.

Just a phase I'm going through?

The Wright-Patterson Air Force Base in Dayton, Ohio, has an amazing sound localization laboratory containing a geodesic sphere, nearly 10 feet across, holding 277 Bose loudspeakers. Listeners zapped by various sounds from all angles indicate where the sounds came from by pointing on a small globe with a special electromagnetic pen. It allows researchers to conduct experiments designed to determine how people can pinpoint sound source location with such stunning accuracy.

A promising new avenue of development for sound localization technology is the microphone array, where systems for extracting voice can be built by setting up a series of closely spaced microphones that pick up different phase patterns. This allows the system to provide better spatial audio and in some cases reconstruct "virtual" microphones to accept or reject certain sounds. In this way, voicing input in noisy environments can sometimes be boosted — a big problem for people with hearing aids.

Harmonizing with harmonics

The basic opening and closing gestures of your vocal folds produce the fundamental frequency (F_0) of phonation. If you were bionic and made of titanium, this is all you would produce. In such a case, your voice would have only a fundamental frequency, and you would sound, well, kind of creepy, like a tuning fork. Fortunately, your fleshy and muscular vocal folds produce more than just a fundamental frequency — they also produce *harmonics,* which are additional flutters timed with the fundamental frequency at numbered intervals. Harmonics are regions of energy at integer multiples of the fundamental frequency. They're properties of the voicing source, not the filter.

Harmonics result whenever an imperfect body — like a rubber band, guitar string, clarinet reed, or vocal fold — vibrates. If you could look at one such cycle, slowed down, with Superman's eyes, you'd see that there's not only a basic (or fundamental) vibration, but also there's a whole series of smaller flutters that are timed with the basic vibration. These vibrations are smaller in amplitude, and (here is the amazing thing) they're spaced in frequency by whole numbers. So, if you're a guy and your fundamental frequency is 130 Hz (also known as your first harmonic), then your second harmonic would be 260 Hz, your third harmonic 390, and so forth. For a female with a higher fundamental frequency, say at 240 Hz, the second harmonic would be 480 Hz, the third 720, and so on. Harmonics are found throughout the speech frequency range (20 to 20,000 Hz). However, there's more energy in the lower frequencies than in the higher because of a 12 dB per octave cutoff.

Extreme harmonics: Phonetics at the edge

A favorite classroom demonstration of mine is to take an enormous strip of rubber from a tire inner tube and stretch it across a phonetics class. Somebody grabs the middle of the inner tube strip and pulls it across to one side of the classroom, everybody ducks, and then the strip is released. As the strip zings back and forth, a few things visibly happen:

- Students can clearly see the fundamental frequency (F_0) of the band as it flies back and forth.
- The band wobbles, showing the harmonics — sub-periodic oscillations that occur at whole number multiples of the fundamental frequency.
- Everyone begins to laugh nervously because (after all) they haven't been hit by the giant, dangerous piece of rubber.
- A few students discreetly call their parents or attorneys.

This is the way of nature — you set up a *simple harmonic series*. Each harmonic series includes a fundamental frequency (or first harmonic) and an array of harmonics that have the relations times 2, times 3, times 4, and so on. Figure 12-6 shows these relations on a vibrating string.

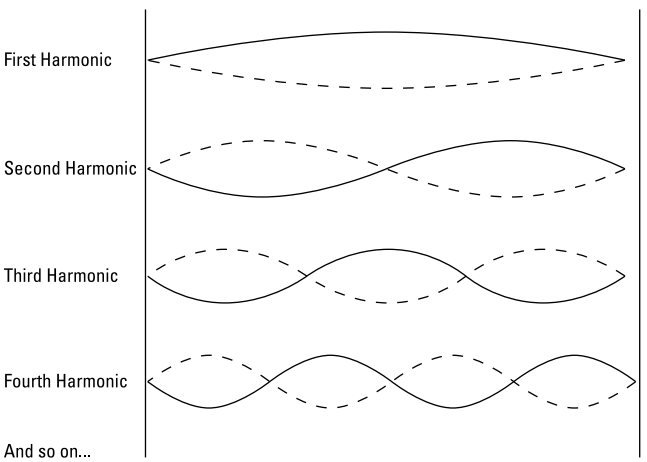


Figure 12-6: Harmonic series on a vibrating string.

Illustration by Wiley, Composition Services Graphics

This spectrum of fundamental frequency plus harmonics gives much of the warmth and richness to the human voice, something in the music world that makes up *timbre* (tone color or tone quality).

Resonating (Ommmm)

Producing voicing is half the story. After you've created a voiced source, you need to shape it. Acoustically, this shaping creates a condition called *resonance,* strengthening of certain aspects of sound and weakening of others. Resonance occurs when a sound source is passed through a structure.

Think about honking your car horn in a tunnel — the sound will carry because the shape of a tunnel boosts it. This kind of resonance occurs as a natural property of physical bodies. Big structures boost low sounds, small structures boost high sounds, and complex-shaped structures may produce different sound qualities.

Think of the shapes of musical instruments in a symphony — most of what you see has to do with resonance. The tube of a saxophone and the bell of a trumpet exist to shape sound, as does the body of a cello.

The parts of your body above the vocal folds (the lips, tongue, jaw, velum, nose, and throat) are able to form complicated passageway shapes that change with time. These shape changes have a cookie-cutter effect on your spectral source, allowing certain frequencies to be boosted and others to be dampened or suppressed. Figure 12-7 shows how this works acoustically during the production of three vowels, /i/, /ɑ/, and /u/.

Imagine that a crazed phonetician somehow places a microphone down at the level of your larynx just as you make each vowel. There would be only a neutral vibratory source (sounding something like an /ə/) for all three. The result would be a spectrum like the one at the bottom of Figure 12-7. Notice that this spectrum has a fundamental frequency and harmonics, as you might expect. When the vocal tract is positioned into different shapes for the three vowels (shown in the middle row of the figure), this has the effect of strengthening certain frequency areas and weakening others. This is resonance. By the time speech finally comes out the mouth, the acoustic picture is complex (as shown in the top of Figure 12-7). You can still see the fundamental frequency and harmonics of the source; however, there are also broad peaks. These are *formants,* labeled F1, F2, and F3.

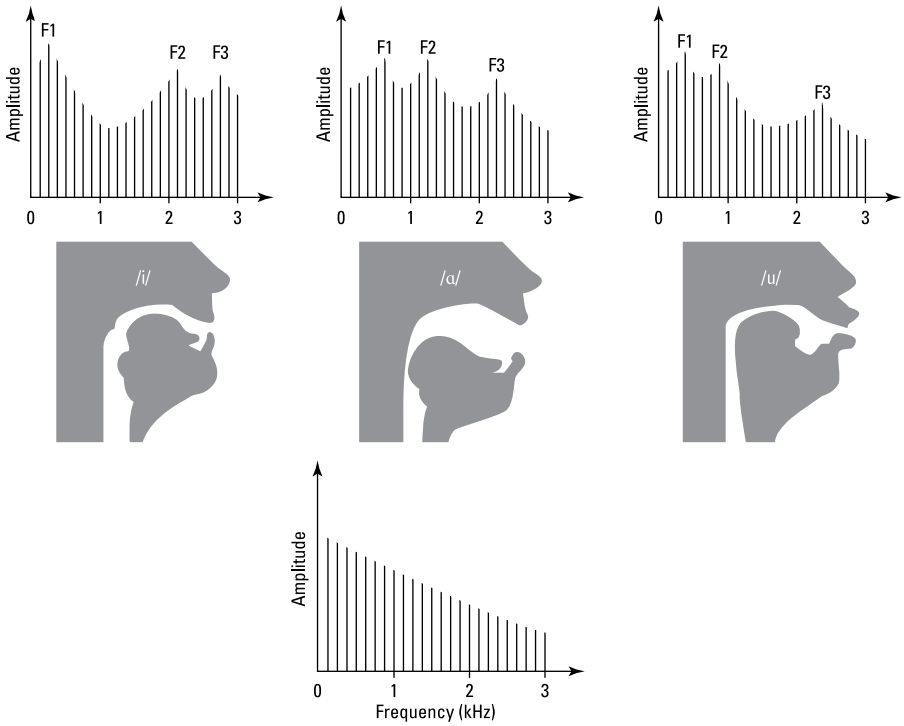


Figure 12-7: Acoustics from line plots of source (bottom), to resonance (middle), to output radiated spectra (top) for /i/ (a), /ɑ/ (b), and /u/ (c).

Illustration by Wiley, Composition Services Graphics

Formalizing formants

These F1, F2, and F3 peaks, called *formant frequencies,* are important acoustic landmarks for vowels and consonants. F1 is the lowest in frequency (shown on the horizontal axis of Figure 12-7, top), F2 is the middle, and F3 is the highest. Phoneticians identify these peaks in speech analysis programs, especially representations called the *sound spectrogram* (one of the most important visual representations of speech sound). Chapter 13 goes into sound spectography in detail. Although usually up to about four to five formants can be seen within the range of most speech analyses, the first three formants are the most important for speech.

Formants provide important information for both vowels and consonants. For vowels, listeners tune in to the relative positions of the first three formant frequencies as cues to typical vowel qualities. Tables 12-1 and 12-2 show values from our laboratory for vowel formant frequencies typical of men and child speakers of American English recorded in Dallas, Texas. Each table has underlined values, which I discuss in greater depth in the "Relating Sound to Mouth" section later in this chapter.

Table 12-1 Mean Formant Frequencies for Men

Vowels	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ʌ/	/ɝ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/
F3	3003	2654	2557	2643	2580	2539	1686	2468	2564	2390	2364	2321
F2	2345	1974	1982	1855	1809	1455	1457	1214	1081	1182	1376	1373
F1	300	445	497	534	694	638	523	754	654	523	426	353

Table 12-2 Mean Formant Frequencies for Children

Vowels	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ʌ/	/ɝ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/
F3	3256	2965	2990	2929	2875	2887	1870	2966	2947	2634	2734	2636
F2	2588	2161	2309	2144	2051	1751	1508	1273	1203	1470	1685	1755
F1	429	522	572	586	836	767	640	688	816	636	516	430

Formant frequency values are commonly used to classify vowels — for instance, in an F1 x F2 plot (refer to Figure 12-8). In this figure, you see that F1 is very similar to what you think of as tongue height and F2 as tongue advancement. This is a famous plot from research done by Gordon Peterson and Harold Barney in 1952 at Bell Laboratories (Murray Hill, New Jersey). It shows that vowels spoken by speakers of American English (shown by the phonetic characters in the ellipses) occupy their own positions in F1 x F2 space — although there is some overlap. For example, /i/ vowels occupy the most upper-left ellipse, while /ɔ/ vowels occupy the most lower-right ellipse. These findings show that tongue height and advancement play an important role in defining the vowels of American English.

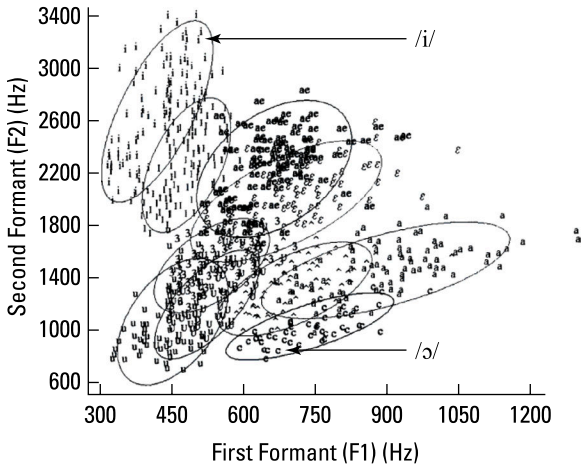


Illustration by Wiley, Composition Services Graphics

Figure 12-8: F2 x F1 plot — American English vowels.

Listening to Hermann von Helmholtz

> "Whoever in the pursuit of science, seeks after immediate practical utility may rest assured that he seeks in vain." —*Academic Discourse* (Heidelberg 1862)

Hermann Ludwig Ferdinand von Helmholtz (1821–1894) was a German physician and physicist who made significant contributions to several areas of modern science. He was a powerhouse thinker, working on fields as diverse as electrodynamics, physiology, psychology, neuroscience, thermodynamics, and mathematics. As if that weren't enough, he was also a philosopher, developing laws of perception, principles of nature, the science of aesthetics, and thoughts about science and civilization.

In acoustics, Helmholtz is remembered for his research into the physics of resonance. In this work, Helmholtz designed beautiful top-shaped vessels (known today as Helmholtz resonators) to study the kind of effect you get when you blow across the top of an empty bottle. The physics of this process turns out to be rather complex and involves the neck of the chamber, the diameter of the opening, and the mass of the air that is forced in.

Based on this work, speech researchers have been able to model some aspects of vowel production as involving Helmholtz-type resonance, especially the part of the mouth behind the tongue. Other aspects of resonance, such as the *F3 rule,* involving the lowering of the third formant in r-colored vowels, behave like a series of coupled Helmholtz resonators interacting.

Formants also provide important information to listeners about consonants. For such clues, formants move — they lengthen, curve, shorten, and in general, keep phoneticians busy for years.

Here are some important points about formants:

- Formants are important information sources for both vowels and consonants.
- Formants are also known as *resonant peaks.*
- Formants are properties of the filter (the vocal tract, throat, nose, and so on), *not* the vocal folds and larynx.
- Formants are typically tracked on a *sound spectrogram.*
- Tracking formants isn't always that easy. In fact, scientists point out formants really can't be *measured,* but are instead *estimated.*

A good way to keep in mind the three most important articulatory (and acoustic) properties of vowels is to keep it funny . . . as in, HAR HAR HAR:

- **H:** *Height* relates inversely to F1.
- **A:** *Advancement* relates to F2.
- **R:** *Rounding* is a function of lip protrusion and lowers all formants through lengthening of the vocal tract by approximately 2 to 2.5 cm.

Relating Sound to Mouth

Don't lose track of how practical and useful the information in this chapter can be to the speech language pathologist, actor, singer, or anyone else who wants to apply acoustic phonetics to his job, practice, or hobby. Because the basic relations between speech movements and speech acoustics are worked out, people can use this information for many useful purposes. For instance, look at these examples:

- Clinicians may be able to determine whether their patients' speech is typical or whether, say, the tongue is excessively fronted or lowered for a given sound.
- An actor or actress may be able to compare his or her impression of an accent with established norms and adjust accordingly.
- A second-language learner can be guided to produce English vowels in various computer games that give feedback based on microphone input.

The physics that cause these F1 to F3 rules are rather interesting and complex. You can think of your vocal tract as a closed tube, a bit like a paper-towel tube closed off at one end. In the human case, the open end is the mouth, and the closed end is the glottis. Such a tube naturally has three prominent formants, as shown in Figure 12-9. It's a nice start, but the cavity resonance of the open mouth modifies these three resonances, and the articulators affect the whole system, which changes the shape of the tube. In this way, the vocal tract is rather like a wine bottle, where the key factor is the shape and size of the bottle itself (the chamber), the length of the neck, and the opening of the bottle (the mouth).

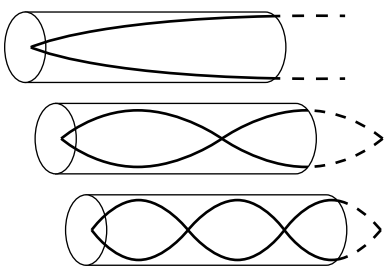


Figure 12-9: Closed-tube model of the vocal tract, showing first three resonances (formants).

Illustration by Wiley, Composition Services Graphics

In the case of your vocal tract (and not the wine bottle), chambers can move and change shape. So sometimes the front part of your chamber is big and the back is small, and other times vice versa. This can make the acoustics all a bit topsy-turvy — fortunately, there are some simple principles one can follow to keep track of everything.

The following sections take a closer look at these three rules and give you some pronunciation exercises to help you understand them. The purpose is to show how formant frequency (acoustic) information can be related to the positions of the tongue, jaw, and lips.

The F1 rule: Tongue height

The *F1 rule* is inversely related to tongue height, and the higher the tongue and jaw, the lower the frequency value of F1. Take a look at the underlined values in Table 12-1 (earlier in the chapter) to see how this works. The vowel /i/ (as in "bee") is a high front vowel. Try saying it again, to be sure. You should feel your tongue at the high front of your mouth. This rule suggests that the F1 values should be relatively low in frequency. If you check Table 12-1 for the average value of adult males, you see it's 300 Hz. Now produce /ɑ/, as in "father." The F1 is 754 Hz, much higher in value. The inverse rule works: The lower the tongue, the higher the F1.

The F2 rule: Tongue fronting

The *F2 rule* states that the more front the tongue is placed, the higher the F2 frequency value. The (underlined) child value for /i/ of 2588 Hz is higher than that of /u/ as in "boot" at 1755 Hz.

The F3 rule: R-coloring

The *F3 rule* is especially important for distinguishing liquid sounds, also known as *r* and *l*. It turns out that every time an r-colored sound is made, F3 decreases in value. (*R-coloring* is when a vowel has an "r"-like quality; check out Chapter 7.) Compare, for instance, the value of male F3 in /ʌ/ as in "bug" and /ɜ˞/ as in "herd." These values are 2539 Hz and 1686 Hz, respectively.

Analyzing things the Fourier way

Joseph Fourier (1768–1830) was a French mathematician and physicist who figured out a clever way to take any function of a variable (such as a wave) and break it down into a series of sine wave multiples. Although Fourier was originally working on the problem of heat flow, his solution was so useful that it's now widely applied in many fields, including phonetics.

Today, any mathematical process that involves decomposing a function into simpler pieces is often called *Fourier analysis* (and the opposite is called *Fourier synthesis*). In phonetics, this typically involves speech analysis and synthesis. The process itself is called a *Fourier transform.* One modern version is called the *Fast Fourier Transform* (or FFT), a version particularly useful on computers.

For phonetic purposes, this means one can take a complex speech signal as input and reduce it to its frequency components (unit sinusoids) by a Fourier analysis, also known as a *harmonic series analysis*. Such analysis can provide useful information about a talker's source characteristics, such as whether the voice is high or low, or healthy or abnormal (breathy or hoarse). Although a group of young students in Japan spent a year with scissors, paper, and glue and managed to work out the Fourier relations of complex waveforms on their own, the math of Fourier analysis is rather involved and is beyond the scope of this book.

The F1–F3 lowering rule: Lip protrusion

The *F1–F3 lowering rule* is perhaps the easiest to understand in terms of its physics. It's like a slide trombone: When the trombonist pushes out the slide, that plumbing gets longer and the sound goes down. It is the same thing with lip protrusion. The effect of protruding your lips is to make your vocal tract (approximately 17 cm long for males and 14 cm for females) about 2.5 cm longer. This will make all the resonant peaks go slightly lower.

Depending on the language, listeners hear this in different ways. For English speakers, it's part of the /u/ and /ʊ/ vowels, such as in the words "suit" and "put." Lip rounding also plays a role in English /ɔ/ and /o/, as in the words "law" and "hope." In languages with phonemic lip rounding, such as French, Swedish, and German, it distinguishes word meaning by lowering sound.

Chapter 13

Reading a Sound Spectrogram

In This Chapter

- Appreciating the importance of the spectrogram
- Decoding clues in spectrogram readouts
- Using your knowledge with clinical cases
- Reading spectrograms that are less than ideal
- Knowing more about noise

The *spectrogram* is the gold standard of acoustic phonetics. These images were originally created by a machine called the *sound spectrograph*, built in the 1940s as part of the World War II military effort. These clunky instruments literally burned images onto specially treated paper. However, software that computes digital spectrograms has replaced this older technology. As a result, you can now make spectrograms on almost any computer or tablet. Although the technology has gotten snappier, you still need to know how to read a spectrogram, and that's where this chapter comes in.

Reading a sound spectrogram is *not* easy. Even highly trained experts can't be shown a spectrogram and immediately tell you what was said, as if they were reading the IPA or the letters of a language. However, with some training, a person can usually interpret spectrograms well for many work purposes. This chapter focuses on making spectrogram reading a bit more comfortable for you.

Grasping How a Spectrogram Is Made

A *spectrogram* takes a short snippet of speech and makes it visual by plotting out formants and other patterns over time. Time is plotted on the horizontal axis, frequency is plotted on the vertical axis, and amplitude is shown in terms of darkness (see Figure 13-1).

Developments in technology have made the production of spectrograms perhaps less exciting than the good ol' days, but far more reliable and useful. Current systems are capable of displaying multiple plots, adjusting the time alignment and frequency ranges, and recording detailed numeric

measurements of the displayed sounds. These advances in technology give phoneticians a detailed picture of the speech being analyzed.

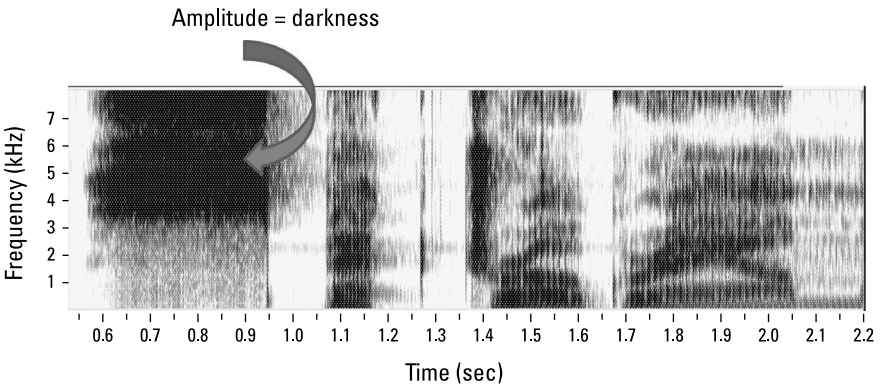


Figure 13-1: A sample spectrogram of the word "spectrogram."

You can easily obtain software for computing spectrograms (and for other useful analyses such as tracking fundamental frequency and amplitude over time) free from the Internet. Two widely used programs are *WaveSurfer* and *Praat* (Dutch for "Speech"). To use these programs, first be sure your computer has a working microphone and speakers. Simply download the software to your computer. You can then access many online tutorials to get started with speech recording, editing, and analysis.

Take a look at Figure 13-2. You can consider the information, shown in a line spectrum, to be a snapshot of speech for a single moment in time. Now, turn this line spectrum sideways and move it over time. Voila! You have a spectrogram. The difference between a line spectrum and a spectrogram is like the difference between a photograph and a movie.

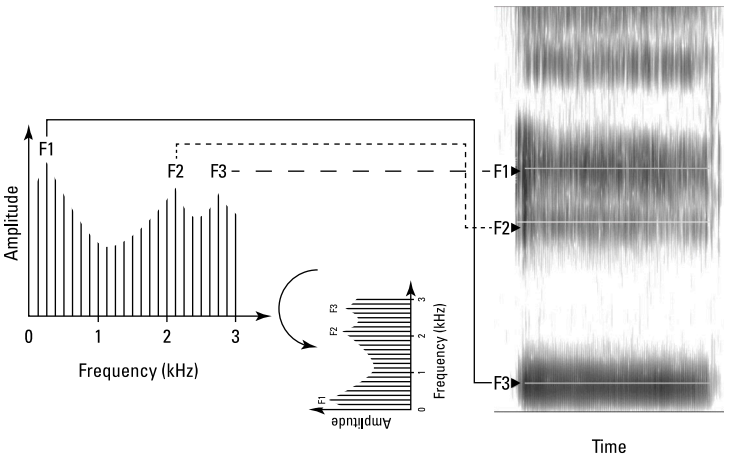


Figure 13-2: Relating the line spectrum to the spectrogram.

Tapping into the history of the spectrogram

Old-fashioned spectrograms used magnetic recording material (either tape or a magnetized metallic bar) and a drum that held a sheet of chemically treated paper marked by an electronic stylus. A person recorded some words or a short phrase into the machine by using a microphone. The spectrograph then sampled energy levels in a small frequency range from the recording and marked those energy levels on the paper. This instrument then analyzed the next frequency range and sampled and marked the energy levels at that point. The process was repeated until the entire desired frequency range was analyzed for that portion of the recording. The finished product was a graphic image of the patterns, including formants, of the acoustical events.

The way these old-fashioned spectrographs worked was by using electronic filters to show formants. Electronic filters act like a coffee filter in that they allow some stuff to pass through (in this case, sound frequencies) while keeping out others. For most spectrograms, the filters are called *band-pass* because they let a band of frequency be captured (for example, between 0 to 300 Hz) and marked by the stylus. By setting the spectrograph to wide band, you could make a broad enough capture to obtain information on formants, the usual display. However, if a person wanted detail on individual harmonics, the filter banks could be set to narrow band. Today, most spectrogram programs simulate wide-band settings.

Old-fashioned spectrographs were really fun because they stimulated all the senses. It took about 80 seconds to make a spectrogram that was about 2.4 seconds in length. For the stylus-marking system to operate, the spectrogram paper first had to be loaded onto a chrome-plated drum, held in place by two circular springs. You then switched the machine on, making the drum spin at dizzying speed. The whole contraption was thrilling for a few reasons:

- Poisonous ozone was emitted from the stylus making contact with the chemical paper.
- If the stylus hit one of the metal springs there would be a short circuit and sparks would fly all over, perhaps starting a fire.
- You still wouldn't really know what kind of pattern you would get (if any) until it was all over.

Don't sell *silence* short while working with spectrograms. Silence plays an important role in speech. It can be a phrase marker that tells the listener important things about when a section of speech is done. Silence can fall between word boundaries (such as in "dog biscuit"). It can be a tiny pause indicating pressure build-up, such as the closure that occurs just before a plosive. Sometimes silence is a pause for breathing, for emotion, or for dramatic effect.

Reading a Basic Spectrogram

Welcome to the world of spectrogram reading. I can see you are new to this, so it's time to establish a few ground rules. You want to read a spectrogram? You had better inspect the axes. Take a look at Figure 13-3.

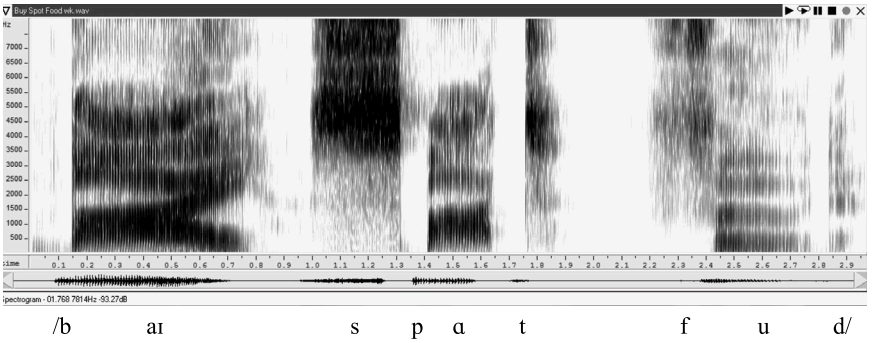


Figure 13-3: Spectrogram of the phrase "Buy Spot food!"

This is the phrase "Buy Spot food!" produced by a male speaker of American English (me). You can assume that Spot is hungry. Actually, I've selected this phrase because it has a nice selection of vowels and consonants to learn. Figure 13-3 is a black-and-white spectrogram, which is fairly common because it can be copied easily. However, most spectrogram programs also offer colored displays in which sections with greater energy light up in hot colors, such as red and yellow.

When reading a spectrogram, you should first distinguish silence versus sound. Where there is sound, a spectrogram marks energy; where there is no sound, it is blank. Look at Figure 13-3 and see if you can find the silence. Look between the words — the large regions of silence are shown in white. In this figure, there is a gap between "Buy" and "Spot" and between "Spot" and "food." I made this spectrogram very easy by recording the words for this speech sample quite distinctly. In ordinary spectrograms of connected speech, distinguishing one word from another isn't so simple.

There are also other shorter gaps of silence, for instance, in the word "Spot" between the /s/ and the /p/. This gap is a *silent gap* that helps distinguish the stop within the cluster. Two other silent regions are found before the final stops at the end of "Spot" (before the /t/) and in food (before the /d/). These are regions of closure before final stop consonant release.

The horizontal axis has a total of about 3,000 milliseconds (or about 3 seconds). If you time yourself saying this same sentence, you'll notice I use a fairly slow, careful rate of speech (*citation form*; as opposed to more usual, informal *connected speech*). In citation form, people tend to be on their best behavior in pronunciation, making all sounds carefully so they can be well understood. I used citation form to make a very clear spectrogram.

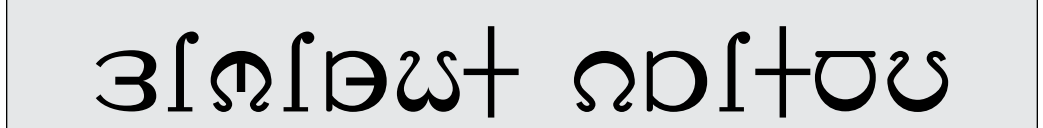
Spy versus spy

A group of researchers at Bell Laboratories in Murray Hill, New Jersey, laid the foundations for the sound spectrograph in the 1940s with programs for "Visible Speech." The idea was to take the auditory speech stream and make it visible. Much of the inspiration for the project was to extend the capabilities of the telephone to create devices for the deaf. Alexander Melville Bell (father to Alexander Graham Bell, the inventor of the telephone) was a phonetician passionately concerned with helping the deaf to communicate. He even invented his own phonetic alphabet (see nearby figure for the phrase "Visible Speech" written in these symbols).

Sound spectrography was also used in World War II. The original development at Bell Labs was carried out as a war project, with spectrograms used for military intelligence purposes.

It appears that a similar program was underway in the Soviet Union. For instance, in the novel *In The First Circle* (written in 1968 by the Russian author Aleksandr Solzhenitsyn), a prisoner camp is described in which technicians and academics were rounded up into gulags in which sound spectrography was done. If phoneticians ran afoul of the authorities, they would be instantly shipped to Siberia. In the novel's plot, a prisoner named Lev Rubin identifies a voice from a recorded phone call and uses spectrograms to eventually arrest a diplomat, who then ends up in the same prison camp.

Most of Solzhenitsyn's work is considered broadly autobiographical. However, there seems to be a bit of a cloak and dagger in the whole business. Solzhenitsyn's writings were an inspiration to many Westerners in exposing the evils of the Soviet system. Others wondered why it was that Solzhenitsyn himself was permitted to write so freely on these topics without being censored by the Soviet State. Some evidence suggests that Solzhenitsyn himself turned informant for the state under the code name "Vetrov," which is what placed him in the comfortable position in the scientific collective where the spectrograms were, instead of being in the harsh gulag in Kazakhstan, where he started out. In real life, Solzhenitsyn tended to side more with people like KGB strongman Vladimir Putin than Russian dissidents such as Andrei Sakharov or Natan Sharansky. Perhaps someday spectrographic evidence will emerge and help settle the issue!



Now, take a look at the vertical axis in Figure 13-3. The frequency ranges from 0 to 7,000 Hz, which is an intermediate range typically used to show both vowels and consonants in spectrograms. To highlight vowels, phoneticians will usually view a lower range (such as to 5,000 Hz), and when sounds with higher frequencies are being inspected (such as fricatives), a higher *y*-axis maximum (for example, up to 10,000 Hz or 20,000 Hz) is sometimes used.

Your next job is to determine whether the sound-containing regions are *voiced* (periodic) or not. A good way to start is to look for energy picked up at the bottom of the frequency scale, which is a band of energy in the very low frequencies, corresponding to the first and second harmonics. For men, it's about 100–150 Hz, for women it's often around 150–250 Hz (with lots of variation between people). If sound is *periodic* (that is, it's due to a regularly vibrating source, such as your vocal folds), a *voice bar* (the dark band running parallel to the very bottom of the spectrogram) will usually be present (although it may be faint or poorly represented, depending on the spectrogram's quality and the talker's fundamental frequency values).

In Figure 13-3, you can see the voice bar at the bottom of "Buy," in the /ɑ/ vowel of "Spot," and in the /ud/ portion of "food." It isn't present for the voiceless sounds, including the /s/ and /t/ sounds of "Spot" and the /f/ of "food."

Visualizing Vowels and Diphthongs

Vowels on a spectrogram can be detected by tracking their steady-state formants over time. A formant appears as a broad, dark band running roughly horizontal with the bottom of the spectrogram page. Some of my more imaginative students have remarked they look like caterpillars (if this helps you, so be it). In that case, you're searching for caterpillars cruising along at different heights, parallel to the spectrogram's horizontal axis.

But how do you know which vowel is which? If you know the talker's gender and accent, then you can compare the center of the formant frequency band with established values for the vowels and diphthongs of English. (If you don't know the gender or accent, your task will be even harder!) Tables 13-1 and 13-2 show formant frequencies for the first (F1), second (F2), and third (F3) vowel formants for common varieties of General American English and British English. Notice that the GAE values are listed separately for men and women, which is relevant because physiological differences in the oral cavity and pharyngeal cavity ratios (and body size differences) between the sexes create different typical values for men and for women. Values for British women weren't available at the time of this writing.

Table 13-1 American English Values

Male	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
F1	342	427	476	580	588	768	652	497	469	378	623	474
F2	2322	2034	2089	1799	1952	1333	997	910	1122	997	1200	1379
F3	3000	2684	2691	2605	2601	2522	2538	2459	2434	2343	2550	1710

Female	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
F1	437	483	536	731	669	936	781	555	519	459	753	523
F2	2761	2365	2530	2058	2349	1551	1136	1035	1225	1105	1426	1588
F3	3372	3053	3047	2979	2972	2815	2824	2828	2827	2735	2933	1929

Table 13-2 British English Values

Male	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
F1	285	356	596	---	748	677	599	449	376	309	722	581
F2	2373	2098	1965	---	1746	1083	891	737	950	939	1236	1381
F3	3088	2696	2636	---	2460	2340	2605	2635	2440	2320	2537	2436

In Figure 13-3, knowing that an American adult male produced "Buy Spot food," you should be able to find the formant frequencies of vowel in the second word shown in the spectrograph.

Figure 13-4 shows the same spectrogram but with additional details about the formant estimates. In this figure, the spectrograph program shows formant frequency values. This figure plots a line in the estimated center frequency of each of the F1, F2, F3, and F4 formants. In old-fashioned spectrograms, a user would have to do this manually, using the eye and a pencil.

How do people do it?

The sounds people listen to in order to hear vowels have been studied for many years, but this problem is still not completely solved. We know that vowel formant frequencies play an important role because synthetic speech can be created from very poorly represented sounds that only contain energy in the F1, F2, and F3 regions and people report that it sounds like vowels. In fact, this is how some of the early (low cost) commercial speech synthesis for toys was done (such as Texas Instrument's popular *Speak and Spell* educational toy in the 1980s).

On the other hand, a professor named Winifred Strange (yes, I know!) and her colleagues at the City University of New York noticed a surprising effect called *silent center* syllable perception. For vowels occurring in CVC syllables (such as "d*ee*d"), the vowel center can be replaced with silence or neutral sounds, thereby removing almost all the vowel formant frequency information, and listeners still can hear the vowel quality rather well. This suggests that there's more to vowel perception than just the steady-state formants. Instead, it seems that information coded in the consonantal portions (often called *dynamic* or *coarticulated* cues) points to vowel information.

The first monophthongal vowel in this phrase is the /ɑ/ in the word "Spot." In Figure 13-4, you can see those values are 724, 1065, and 2571 Hz. These map quite closely to the formant values for the male American /ɑ/ shown (768, 1333, and 2522 Hz).

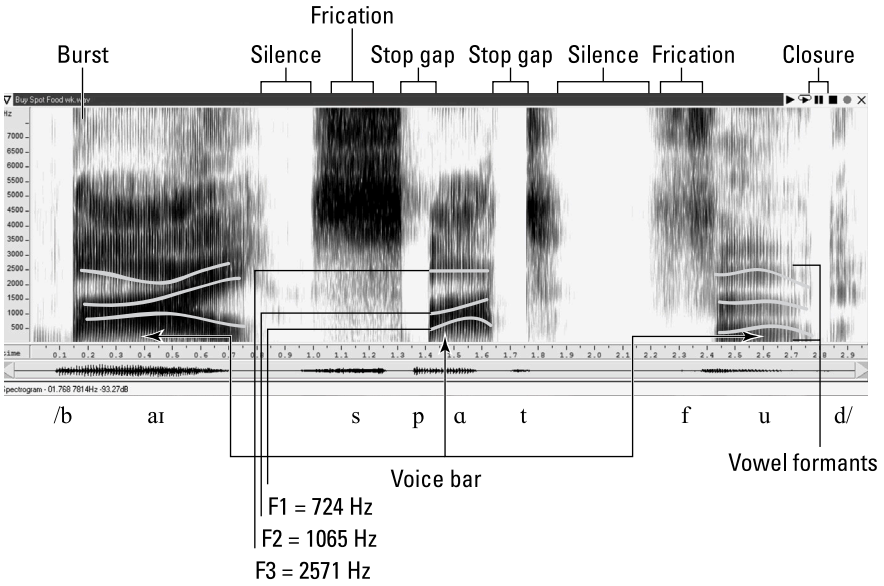


Figure 13-4: An annotated spectrogram.

Next, examine the /u/ of "food." In Figure 13-4, the F1, F2, and F3 values are estimated in the same fashion. These are 312, 1288, and 2318 Hz. You can see that these measurements match closely to the /u/ values for the GAE male talkers in Table 13-1 (378, 997, and 2343 Hz). My F2 is a bit higher, perhaps because I'm from California and it seems to be a dialectal issue in California, where "u" vowels begin rather /i/-like. Overall, the system works.

Vowels that behave this way are traditionally called *steady state* because they maintain rather constant formant frequency values over time. Another way of putting it is that they have relatively little *vowel inherent spectral change (VISC).*

In contrast, the General American English diphthongs (/aɪ/, /aʊ/, and /ɔɪ/) perceptually shift from one sound quality to another. Acoustically, these diphthongs show relatively large patterns of formant frequency shift over time, as in "buy" shown in Figure 13-5. Spectrograms of /aɪ/, /aʊ/, and /ɔɪ/ are shown in Figure 13-5, for comparison.

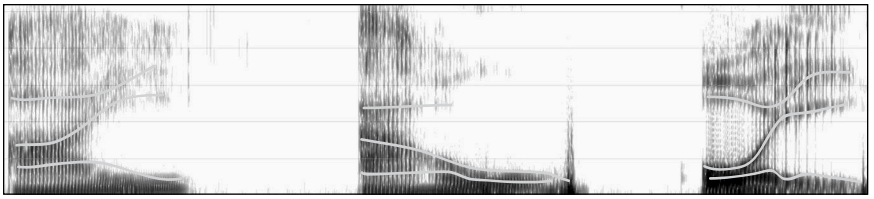


Figure 13-5: Spectrograms of /aɪ/, /aʊ/, and /ɔɪ/.

Diphthongs provide an excellent opportunity to review the rules mapping formant frequencies to physiology (refer to Chapter 12). For instance, in /aɪ/ you see that according to the F1 rule when the tongue is low for the /a/ of /aɪ/, F1 is high. However, F1 drops when the tongue raises for the high vowel /ɪ/ at the end of the diphthong. Conversely, /a/ is a central vowel, while /ɪ/ is a front vowel. According to the F2 rule, F2 should increase as one moves across the diphthong (and indeed this is the case).

Checking Clues for Consonants

Consonants are different beasts than vowels. Vowels are voiced and relatively long events. You make vowels by positioning the tongue freely in the mouth. That is, the tongue doesn't need to touch or rub anywhere. Consonants can be long in duration (as in fricatives) or short and fast (like stops). Consonants involve precise positioning of the tongue, including movement against other articulators.

Identifying consonants on spectrograms involves a fair bit of detective work because you must go after several clues. Your first clue is the manner of articulation. Recall that there are stops, fricatives, affricates, approximants, and nasals. In these sections, I show you some of each. Later in the chapter, after you know what each of these manner types look like on the spectrogram, I explain the place of articulation (labial, alveolar, velar, and so on) for stop consonants, a slightly more challenging task in spectrogram reading.

Stops (plosives)

Stop consonants can be identified on spectrograms because of their brevity: they're rapid events marked by a burst and transition. Say "pa ta ka" and "ba da ga." Feel the burst of each initial consonantal event. Now look at the spectrograms in Figure 13-6. Notice that each has a thin and tall pencil-like spike where the burst of noise has shot up and down the frequency range. As you might expect, the voiced stops have a voice bar underneath, and for the voiceless cases, there aren't voice bars.

Stop consonants look rather different at the end of a syllable. First, of course, the transitions are pointing in the opposite direction than when the consonant is at the beginning of the syllable. Also, as you saw in Figure 13-3 with the final consonants in "Spo*t*" and "foo*d*," there is a silent closure before the final release. Figure 13-7 shows two more examples, "pat" and "pad," with important sections labeled.

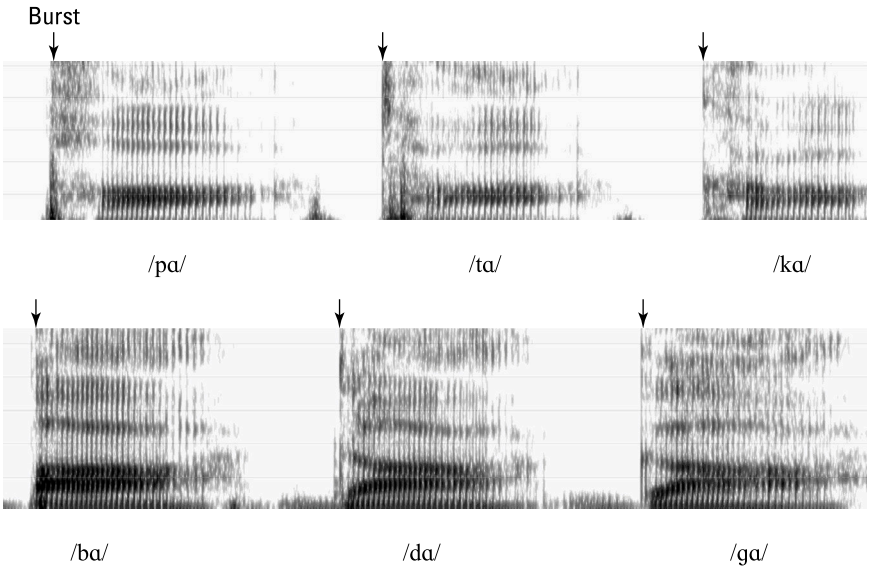


Figure 13-6: The spectograms of /pɑ/, /tɑ/, /kɑ/ (top) of /bɑ/, /dɑ/, and /gɑ/ (bottom).

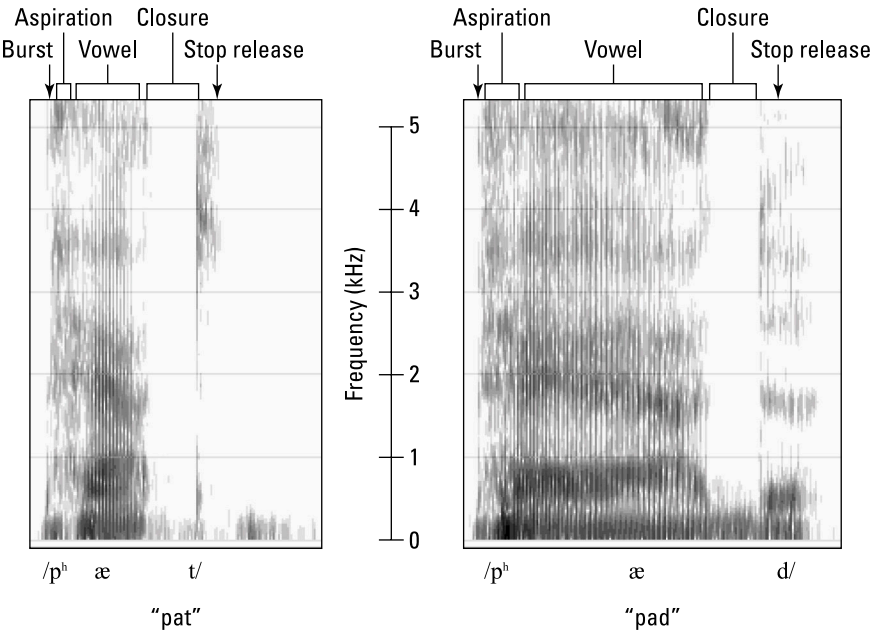
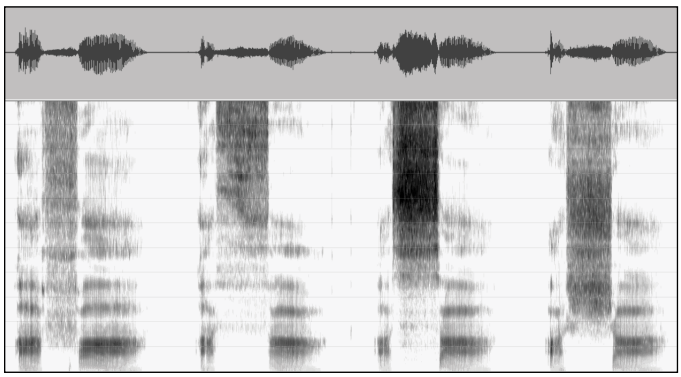


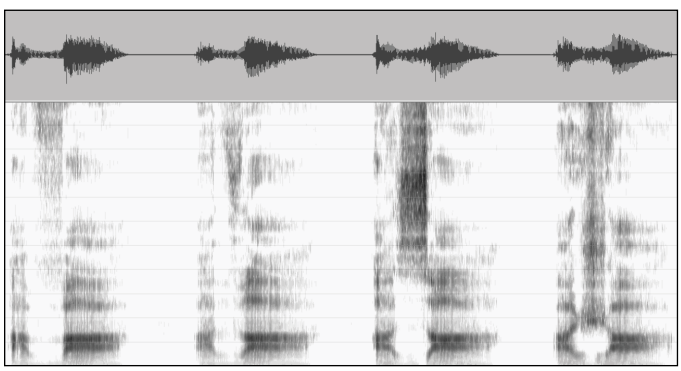
Figure 13-7: Spectrogram of "pat" and "pad."

Fricative findings

Noise (friction) shows up in spectrograms as darkness (intensity marking) across a wide frequency section. Figure 13-8 shows the voiced and voiceless fricatives of English in vowel, consonant, vowel (VCV) contexts.



/ɑfɑ/ /ɑθɑ/ /ɑsɑ/ /ɑʃɑ/



/ɑvɑ/ /ɑðɑ/ /ɑzɑ/ /ɑʒɑ/

Figure 13-8: The spectrograms of GAE fricatives in VCV contexts.

Here is a list of important fricative points to remember:

- Fricatives are fairly long. Their durations are clearly longer than stop consonants.
- The voice bar can be a good cue for telling the voiced from the voiceless.
- The energy distribution (spread) of the different fricatives isn't the same. Some are darker in higher frequency regions, some in lower regions.

- /s/ and /ʃ/ are produced with strong airflow (*sibilants*).
- /f/, /v/, /ð/, and /θ/ are produced with weak airflow (*non-sibilants*).

Energy spread is an especially good clue to fricative identity. If you listen to /s/ and /ʃ/, you hear that these are strong and hissy because they're made by sharply blowing air against the teeth, in addition to the oral constriction. Compare /s/ and /ʃ/ (the strong fricatives, or *sibilants*) with /f/, /v/, /ð/, and /θ/. This second group should sound weaker because they don't involve such an obstacle.

Tuning in to the sibilants, you can also hear that /s/ sounds *higher* than /ʃ/. This shows up on the spectrogram with /s/ having more darkness at a higher frequency than does /ʃ/. In general, /s/ and /z/ have maximum noise energy, centering about 4000 Hz. For /ʃ/ and /ʒ/, the energy usually begins around 2500 Hz.

Okay, the strong fricatives are out of the way, so you can now work over the weaklings (non-sibilants). A characteristic of this whole group is they may not last as long as /s/ and /ʃ/. Because of this (and because of their weak friction) they may sometimes look like stops. Don't let them get away with it: Check out the lineup in Figure 13-8.

The fricatives /f/ and /v/ are the strongest of the weaklings. They can show up on the spectrogram as a triangular region of frication. In most cases there is strong energy at or around 1200 Hz. The fricative /θ/ can take two forms:

- A burst-like form more common at syllable-initial position
- A more fricative-like pattern at the end of a syllable (shown in Figure 13-8)

It can sometimes be accompanied by low-frequency energy. However, its frication is usually concentrated above 3000 Hz.

The phoneme /ð/ is the wimpiest of all the fricatives; it can almost vanish in rapid speech, although unfortunately this sound occurs in many common function words in English (*the, that, then, there,* and so on). When observable, /ð/ may contain voiced energy at 1500 and 2500 Hz, as well as some higher-frequency energy.

Affricates

English has two affricates, /tʃ/ and /dʒ/. These have an abrupt (alveolar) beginning, marked with a burst and transition, followed by energy in an alveolar locus (approximately 1800 Hz). This quickly transitions into a palato-alveolar fricative. Old spectrogram hands suggest a trick for pulling out affricate suspects from the lineup: Sometimes there's a bulge in the lower

frequency portions of the fricative part. The plosive component is detectable as a single vertical spike just to the left of the frication portion of the phoneme! Check out Figure 13-9 for such evidence.

Approximants

Approximants have more gradual transitions than those of stops, as seen in Figure 13-10. This spectrogram shows the approximants found in GAE, including /w/ and /j/, two approximants also called *glides.* They have this name because these consonants smoothly blend into the vowel next to them. They also have less energy than that of a vowel. A time-honored phonetician's trick for spotting /j/ is to look for "X marks the spot" where F2 and F3 almost collide before going their merry ways. Because the constriction for /j/ is so narrow, this phoneme is often marked by frication as well as voicing.

The sounds /ɹ/ and /l/ are fun because of the unique tongue shapes involved. Taken together, these two approximants are called *liquids* because of the way these sounds affected the timing of the classical Greek language. The "r" sounds (*rhotics*) are a particularly scandalous bunch. Literally. They may involve a bunched tongue, as in some forms of American English, a *retroflex* gesture (bringing the sides of the blade curled up to the alveolar ridge and the back tongue sides into contact with the molars), uvular fricatives (such as in French or Hebrew), taps, or trills. Looking at the American English /ɹ/ in Figure 13-10, the main acoustic characteristic becomes clear: A sharp drop in F3.

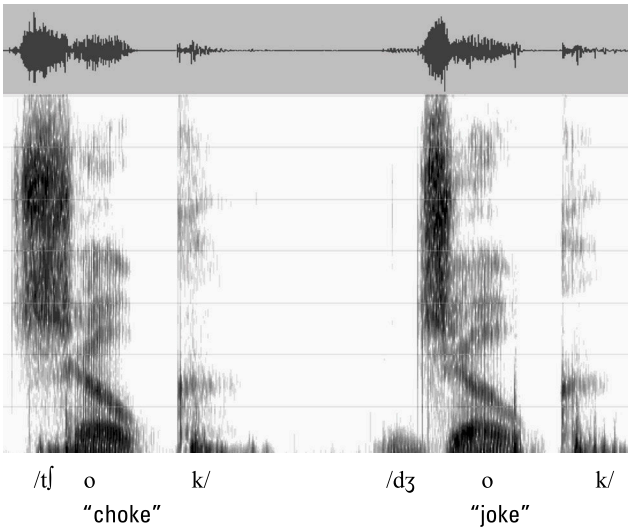


Figure 13-9: The spectrograms of /tʃ/ and /dʒ/.

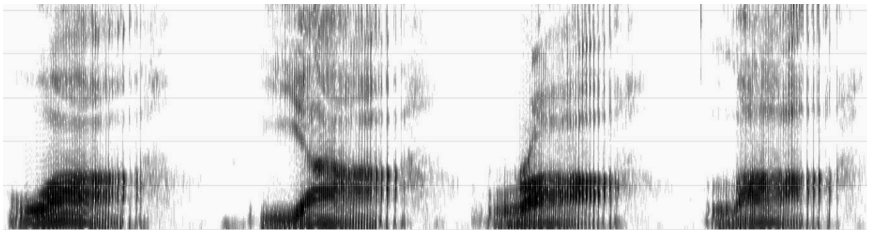


Figure 13-10: The spectrograms of approximants /wɑ/, /jɑ/, /ɹɑ/, and /lɑ/.

The lateral approximant /l/ creates a side-swiped situation in the oral cavity. In a typical /l/ production, the tongue tip is placed on the alveolar ridge and the sides are in the usual position (or slightly raised), with air escaping around the sides. This causes something called *anti-resonance* at 1500 Hz, which you can see as a fading out of energy in that spectrogram zone. Anti-resonance is an intensity minimum or zero.

Spectrograms that contain /l/ consonants can show much variability. For example, before a vowel F3 may drop or stay even, while F2 rises, giving the phoneme a forked appearance. Following a vowel, /l/ may be signaled by the merging of F2 with F1 near or below 1000 Hz, with F3 moving up toward 3000 Hz, leaving a hole in the normal F2 side-swiped by /l/, acoustically.

Nasals

Imagine you entered a futuristic world where a nasty government went around spying on everyone by using voice detectors to snatch all kinds of personal information from people. How could you escape detection? The first thing I would do is change my name to something like "Norman M. Nominglan." That is, something laden with nasals. This is because nasals are some of the most difficult sounds for phoneticians to model and interpret. They're tough to read on a spectrogram and tend to make speech recognizers crash all over the place. Go nasal and fly under the radar.

English has three nasal stop consonants, bilabial /m/, alveolar /n/, and velar /ŋ/. They're produced by three different sites of oral constriction, and by opening of the velar port to allow air to escape through the nasal passageway. Opening the nasal port adds further complexity to an already complicated acoustic situation in the oral cavity. As in the case of /l/, nasal sounds have anti-resonances (or zeros), which can show up in spectrograms. To help you track down anyone named "Norman M. Nominglan," here are some important clues:

- Nasal consonants are voiced events, but they have lower amplitudes than vowels or approximants. Nasals therefore appear fainter than surrounding non-nasal sounds.

- There may be a characteristic *nasal murmur* (sound that occurs just after oral closure) at 250 Hz, near F1.
- If nasals are at the start or end of a syllable, F1 may be the only visible formant.
- Nasal stops (like other plosives) have an optional release.
- F2 is the best clue for place of articulation. F2 moves toward the following target values:
  - /m/ for bilabials 900 to 1400 Hz.
  - /n/ for alveolars 1650 to 1800 Hz.
  - /ŋ/ for velars 1900 to 2000 Hz.

Check out the suspects in Figure 13-11.

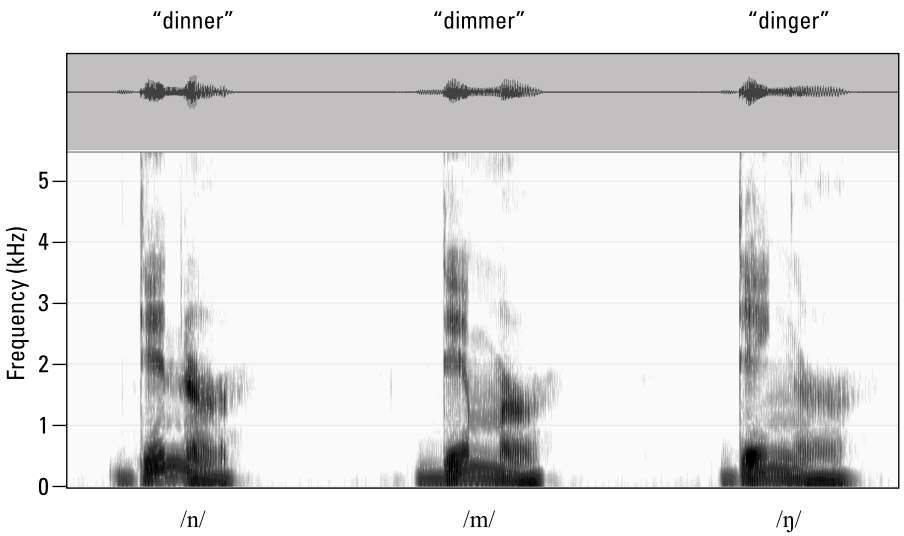


Figure 13-11: The spectrograms of /n/, /m/, and /ŋ/.

Formant frequency transitions

An important basis for tracking consonant place of articulation in spectrograms is the *formant frequency transition,* a region of rapid formant movement or change. Formant frequency transitions are fascinating regions of speech with many implications for speech science and psychology. A typical formant frequency transition is shown in Figure 13-12.

If a regular formant looks like a fuzzy caterpillar, then I suppose a formant frequency transition looks more like a tapered caterpillar (or one wearing styling gel). This is because the transition begins with low intensity and a narrow bandwidth, gradually expanding into the steady state portion of the sound.

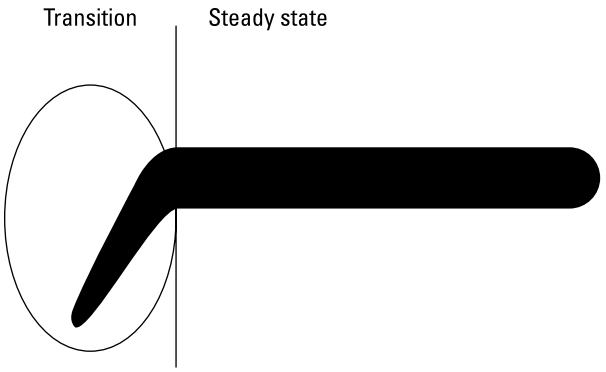


Figure 13-12: A typical formant frequency transition.

Here's how it works.

- **F1:** Think about what your tongue does when you say the syllable "da." Your tongue moves quickly down (and back) from the alveolar ridge. Following the inverse rule for F1, it means that F1 rises. Because you're moving into the vowel, the amplitude also gets larger.
- **F2:** These transitions are a bit trickier. For stop consonants, transitions, F2 frequency transitions are important cues for place of articulation. Figure 13-13 shows typical F1 and F2 patterns for the *nonce* (nonsense) syllables /bɑ/, /dɑ/, and /gɑ/. Notice that these transition regions start from different frequency regions and seem to have different slopes. For the labial, the transition starts at approximately 720 Hz and has a rising slope. The alveolar stop, /d/, starts around 1700–1800 Hz and is relatively flat. The velar stop, /g/, begins relatively high, with a falling slope. A common pattern also seen for velars is a *pinching* together of F2 and F3, where F2 points relatively high up and F3 seems to point to about the frequency region.

  Phoneticians use these stop-consonant regions, called the *locus*, to help identify place of articulation in stop consonants. The physics behind these locus frequencies is complex (and a bit beyond the scope of this book). However, in general they result from interactions of the front and back cavity resonances.

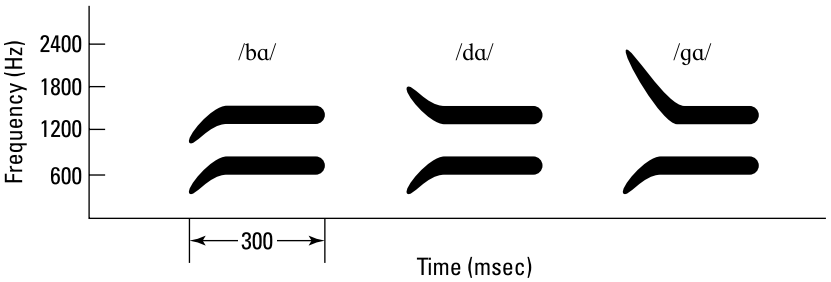


Figure 13-13: Stylized F1 and F2 patterns for /bɑ/, /dɑ/, and /gɑ/.

Hey! Some of my spectrograms don't look like the ones in the book

The examples in this book (and in others) serve as useful starting points. However, individual talkers will show considerable variability with respect to the patterns produced for speech sounds. Variability is found both within a given talker (*intra*-talker) and between talkers (*inter*-talker). Beginning with intra-talker variability, if you make spectrograms of yourself you'll notice that you speak differently if you're talking loudly or quietly, with different types of emotion (affect), when something is said quickly or slowly, and even if your body position is different (for example, sitting up versus lying on your back).

An even more whopping source of variability is found between talkers. As you go from person to person, sex, accent, speaking style, and individual physiology can all enter in to produce differences. This is why, for example, frequency norms for formants are an approximation only. It is also why patterns that commonly show up in clear speech (for example, in citation form) may not appear at all in more casual speech.

Finally, spectrograms tend to be sexist as all get out. They discriminate against women and children because such talkers happen to have higher fundamental frequencies of phonation. As is described later in the chapter, this creates a problem in estimating formants. Without formants, one can't get very far with a spectrogram. Sorry, ladies and kiddies. Nothing personal!

These rapidly changing sections of the speech signal are integrated by people's perceptual systems in a smooth, seamless fashion. For instance, imagine you create a synthetic syllable on a computer ("da") and then artificially chop out just the formant frequency transitions (for example, just for the "d"). If you play this section, it won't sound like a "d"; it will instead just sound like a click or a stick hitting a table. That is, there is not much speech value in formant frequency transitions alone. They must be fused with the neighboring steady state portion in order to sound speech-like.

Spotting the Harder Sounds

A few sounds on the spectrogram may have escaped your detection. These sounds typically include /h/, glottal stop, and tap. Here are some clues for finding them.

Aspirates, glottal stops, and taps

The phoneme /h/ has been living a life of deceit. Oh, the treachery! Technically, /h/ is considered a glottal fricative, produced by creating friction at the glottis. It is unvoiced. This is all very well and fine, except for the fact that when

phoneticians actually investigated the amount of turbulence at the glottis during the production of most /h/ consonants, they discovered, there is almost no friction at the glottis for this sound.

In other words, /h/ is scandalously misclassified. Some phoneticians view it as a signal of partial devoicing for the onset of a syllable. Others call it an *aspirate,* as in the diacritic for aspiration [^{h}]. You can observe a spectrogram of /h/ in Figure 13-14. One nice thing about /h/ is that it's very good at fleshing out any formant frequencies of nearby (flanking) vowels. They'll run right through it.

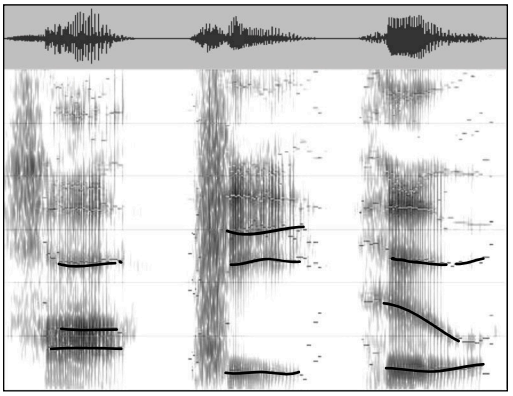


Figure 13-14: The spectrograms of /h/: /hɑ/, /hi/, and /hu/.

Ready for the big time?

Some great websites can keep your spectrogram reading abilities sharp. Robert Hagiwara, a professor at the University of Manitoba, has for many years sponsored a Mystery Spectrogram Webzone at `http://home.cc.umanitoba.ca/~robh/howto.html#formants`. He posts a mystery spectrogram and challenges you to decode it. He also posts past challenges with solutions. The site includes spectrogram examples and a tutorial. Highly recommended!

Another superb spectrogram reading site is maintained by Steve Winters, a professor at the University of Calgary. This webpage ("Spectrogram Reading, for Fun and Profit") features clues, such as "Cult classic TV catch phrase." Check out `https://webdisk.ucalgary.ca/~swinters/public_html/ling441/spectrograms.html`.

You may now turn, with relief, to another sound made at the glottis that is much simpler, the glottal stop. This is marked by silence. Clean silence. And relatively long silence. For instance, look at "uh oh" in Figure 13-15. The silent interval of glottal stop is relatively long.

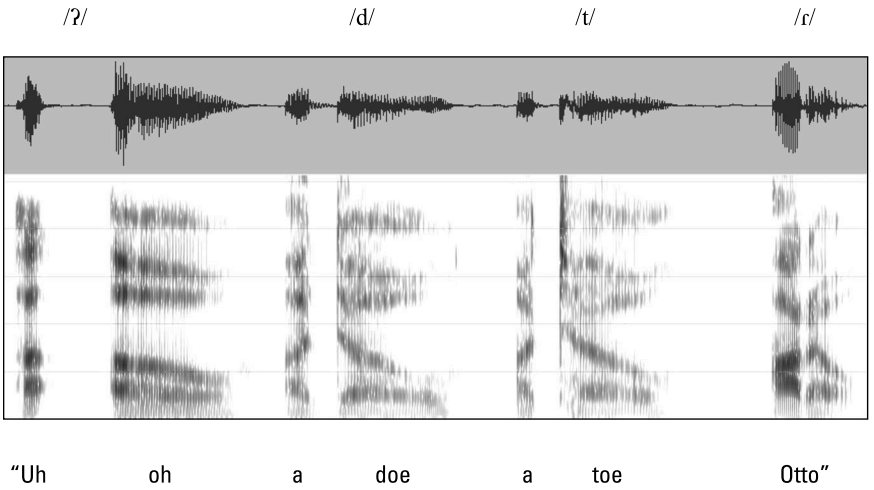


Figure 13-15: The spectrograms of /ʔ/ and /ɾ/ compared with /d/ and /t/.

The glottal stop may be contrasted with the alveolar tap, /ɾ/, a very short, voiced event. In American English, this is not a phoneme that stands by itself. Rather it is an allophone of the phonemes /t/ and /d/. Contrast "a doe," "a toe," and "Otto" (GAE accent) in Figure 13-15. Here are some hints for spotting taps:

- A tap is among the shortest phonemes in English — as short as two or three pitch periods.
- The English tap usually has an alveolar locus (around 1800 and 2800 Hz).
- There is often a mini-plosion just before the resumption of the full vowel after the tap. The mini-plosion occurs when the tongue leaves the alveolar ridge.

Cluing In on the Clinical: Displaying Key Patterns in Spectrograms

Spectrograms can be an important part of a clinician's tool chest for understanding the speech of adult neurogenic patients, as well as children with speech disorders. Chapter 19 gives you added practice and examples useful for transcribing the speech of these individuals.

A spectrogram can handily reveal the speech errors of communication-impaired talkers. This section gives you some examples of speech produced by individuals with error-prone speech, compared with healthy adult talkers, for reference. The first two communication-impaired talkers are monolingual speakers of GAE, the last is a speaker of British English.

- Female with Broca's aphasia and AOS (Apraxia of speech)
- Female with ALS (Amyotrophic lateral sclerosis)
- Male with cerebral palsy (spastic dysarthia)

In Figure 13-16, the subject describes a story about a woman being happy because she found her wallet. The intended utterance is "And she was relieved." There is syllable segregation — the whole phrase takes pretty long (try it yourself; it probably won't take you 3 seconds). There are pauses after each syllable (as seen in the white in the spectrogram). I am sure you don't do this either. There is no voicing in the /z/ of /wəz/ (note the missing voice bar) and the final consonant is also missing in the ending of "relie*ved,*" which comes out as a type of /f/, heard as "relie*f*."

Dysarthria occurs in more than 80 percent of ALS patients and may cause major disability. Loss of communication can prevent these patients from participating in many activities and can reduce the quality of life. Dysarthria is often a first symptom in ALS and can be important in diagnosis.

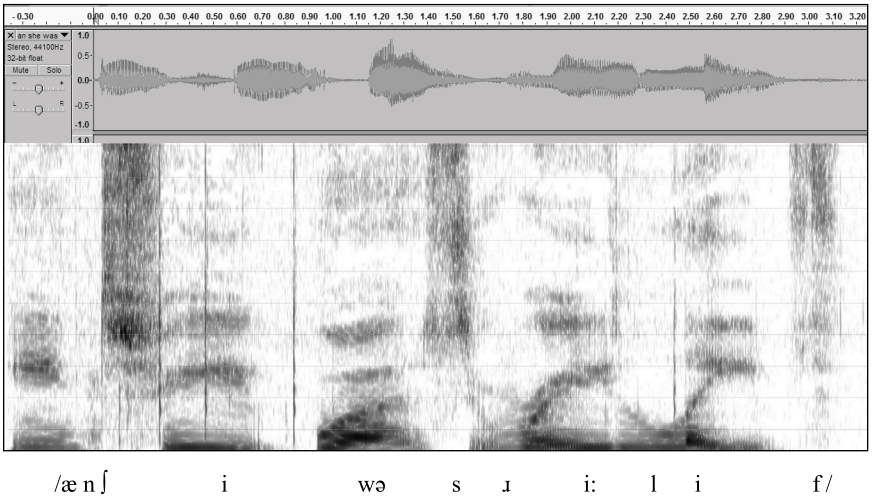


Figure 13-16: The spectrogram of an individual with BA and AOS showing syllable prolongation.

There are many ways ALS speech can be noted in a spectrogram. Figure 13-17 gives one common example. Look at the syllables /bib/, /beb/, and /baeb/ produced by an individual with ALS having moderate-to-severe dysarthria (66 percent intelligibility), compared with those of an age-matched control talker. You will notice a couple of things:

- The productions by the individual with ALS are slightly longer and more variable.
- Whereas the healthy talker has nice sharp bursts (viewable as pencil-like spikes going up and down the page), the productions of the ALS talker have none. This is graphic evidence of why she sounds like she does: instead of sounding like a clear /b/, the oral stops sound muted.
- The broadened formant bandwidths and reduced formant amplitudes suggest abnormally high nasalization.

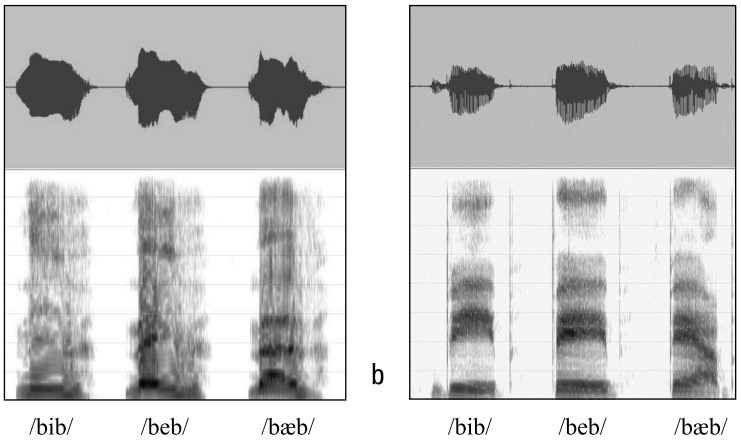


Figure 13-17: The spectrograms of ALS speech (a) and healthy speech (b).

Take the spectrogram to clinic

A group at Portland State University has launched a project called Spectrogram for Speech, designed to promote the use of visual feedback through spectrographic displays for treating a variety of speech sound disorders. At `http://www.spectrogramsforspeech.com`, Jess Leigh Bullock guides the user through the background and theory of spectrogram usage for a variety of applications in speech language pathology. This includes a tutorial in obtaining, running, and interpreting data from WaveSurfer.

People with cerebral palsy (CP) commonly have dysarthria. The speech problems associated with CP are poor respiratory control, laryngeal, and velopharyngeal dysfunction, as well as oral articulation disorders that are due to restricted movement in the oral-facial muscles. You can find more information on CP and dysarthria in Chapter 19.

The next spectrograms highlight spastic dysarthria in a talker with CP. Speech problems include weakness, limited range of motion, and slowness of movement. In this spectrogram (Figure 13-18), you can see evidence of issues stemming from poor respiratory control and timing. In the first attempt of the word "actually," the pattern shows a breathy, formant-marked *vocoid* (sound made with an open oral cavity) with an /æ/-like value, then the consonant /ʧ/, followed by a /d/-like burst, slightly later. There is then an intake of air and a rapid utterance of "I actually just" in 760 ms. This time, the final /t/ isn't realized.

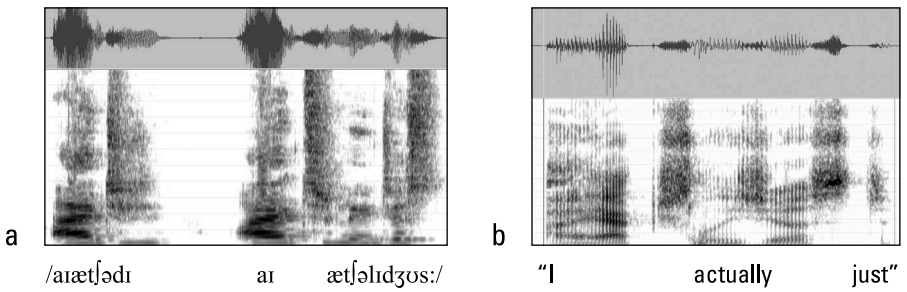


Figure 13-18: The spectrograms of spastic dysarthria in cerebral palsy (a), compared with healthy speech (b).

If you compare this with the same thing said (rapidly) by a control speaker, notice that formant patterns are nevertheless relatively distinct in the spectrogram of the healthy talker, particularly formant frequency transitions and bursts. There is formant movement in and out of the /l/. There is a /k/ burst for the word "actually" and the final /t/ of "just".

Working With the Tough Cases

Certain speaker- and environment-dependent conditions can make the task even more difficult for reading spectrograms. These sections take a closer look at these tough cases and give you some suggestions about how to handle them.

Women and children

Tutorials on spectrogram reading generally try to make things easy by presenting clear examples from male speakers and by using citation forms of speech. There's nothing wrong with that! Until, of course, you must analyze your first case of a child or female with a high fundamental frequency. At this point, you may see your first case of spectrogram failure, where formants simply won't appear, as expected. Take a look at Figure 13-19. This figure shows a man, woman, and 5-year old child each saying the word "heed" (/hid/ in IPA) and having the fundamental frequencies 130 Hz, 280 Hz, and 340 Hz, respectively. Notice that the formants in the spectrograms of speech produced by the man and the woman are relatively easy to spot, while those of the young child are fuzzy (F1 and F2) or missing entirely (F3).

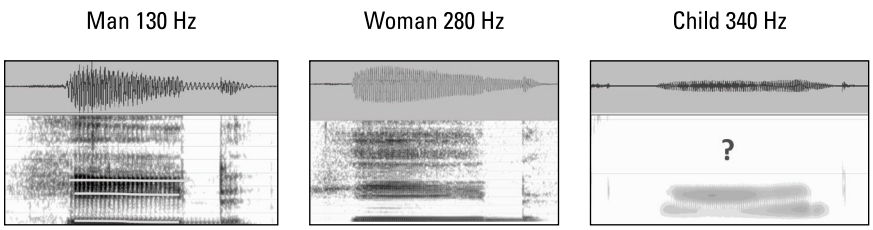


Figure 13-19: The spectrograms of /hid/ by a man, woman, and child with F_0s indicated.

The reason for the decreasing clarity is a problem called *spectral sketching*, a problem of widely spaced harmonics in cases of high fundamental frequencies. Recall that the spectrograph's job is to find formants. It does this either by using bandwidth filters, which is old school, or by newer methods, such as fast Fourier transform (FFT) and linear predictive coding (LPC) algorithms. If, however, a talker has a high voice, this results in relatively few harmonics over a given frequency band. As a result, there isn't much energy for the machine or program to work with. The spectrum that results is sketchy; the system tends to resolve harmonics, instead of formants as it should.

Figure 13-20 shows a male vocal tract with a deep voice and its harmonics compared with a child vocal tract and its harmonics. Figure 13-20a and 13-20b show a snapshot of the energy taken at an instant in time. There is

more acoustic information present in the male's voice that can be used to estimate the broad (formant) peaks. However, in the child's voice, the system can't be sure whether the peaks represent true formants or individual harmonics. There is just not enough energy there.

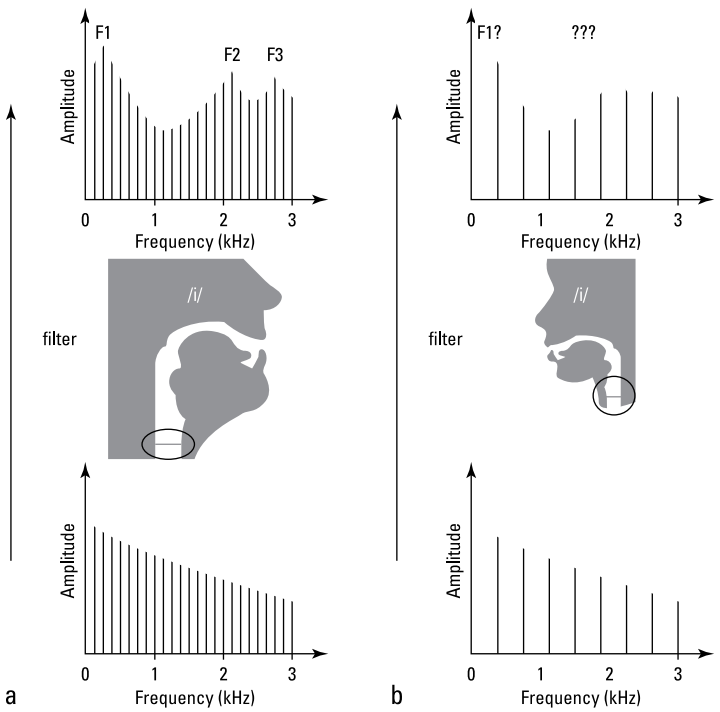


Figure 13-20: A male's (a) and a child's vocal tracts (b) with line spectra input (below) and the results of vocal tract filtering (above).

Speech in a noisy environment

Another challenge with many applications, from working with the deaf, to forensics, to military uses, is detecting a meaningful speech signal from a noisy environment.

Noise can be defined as unwanted sound. It can be regular, such as a hum (electric lights) or buzz (refrigerator, air conditioner), or random-appearing and irregular sound (traffic sounds, cafeteria noise).

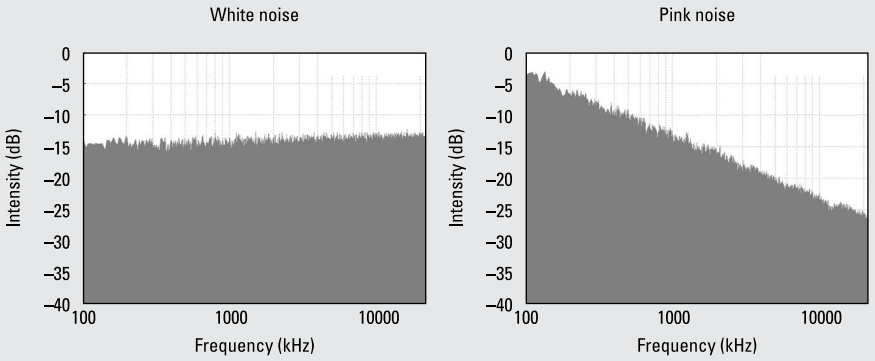
What color is your noise?

To make noise from a computer, you generate it from a random signal. Because of this, it can have different properties associated with this randomness and its relation to the output spectrum. For this reason, acousticians, engineers, and physicists use spectral density (the power distribution in the frequency spectrum) to describe different types of noise. Noise with different spectral density is given color terminology, with different types named after different colors, including pink noise, blue noise, violet noise, and grey noise. (See the nearby figure for examples of white and pink noise.)

Audiologists and phoneticians sometimes use white noise and pink noise for various types of experiments and applications:

- *White noise* (named by analogy to white light) is a mix of sound waves with equal power with a broad frequency bandwidth. It has a flat power spectrum. One application is its use to help people with *tinnitus* (ringing in the ears) cope with their symptoms. White noise masking systems can give relief to some tinnitus sufferers.
- *Pink noise* is acoustical energy distributed evenly by octave throughout the range of human hearing (approximately 20 Hz to 20 kHz). Most people hear pink noise as having the same loudness at all frequencies. The total sound power of pink noise in each octave is the same as the total sound power in the octave immediately above or below it.

For spectrogram reading, one principle about noise is simple and effective: The less, the better! If noise is too loud and broad, it can swallow up any patterns in your spectrogram. Do any recordings for phonetics in the quietest setting you can find. The best rooms have carpets, not wooden floors. No TV or radio in the background. And definitely no espresso machines!



Lombard effect

People naturally increase the loudness of their voices when they enter a noisy room to make their voices clearer. This is called the *Lombard effect* (named after the French otolaryngologist, Etienne Lombard). What is surprising is that people do more than simply increase their volume. They also typically raise their F_0, make their vowels longer, change the tilt of their output spectrum, alter their formant frequencies, and stretch out *content words* (such as nouns and verbs) longer than *function words* (such as "the," "or," and "a").

Incidentally, humans aren't alone. Animals that have been found to alter their voices in the Lombard way are budgies, cats, chickens, marmosets, cotton top tamarins, nightingales, quail, rhesus macaques, squirrel monkeys, and zebra finches.

Cocktail party effect

The *cocktail party effect* is quite different than the Lombard effect (see the preceding section). It's a measure of *selective attention*, how people can focus on a single conversation in a noisy room while "tuning out" all others. People are extremely good at this — much better than machines. To test this for yourself, try recording a friend during conversation in a noisy room and later play the recording back to see if you can understand anything. You may be surprised at how difficult it is to hear on the recording what was so easy to detect "live" and in person in the room.

Such focused attention requires processing of the phase of speech waveform, resolved by the use of *binaural hearing* (involving both ears). Chapter 2 includes information on the phase of speech waveforms. In a practical sense, some people will resort to the *better ear effect,* in which one ear is cocked toward the conversation and farther from the party noise, as a strategy.

How people attend cognitively to the incoming signal is less well understood. Early models suggested that the brain could sharply *filter* out certain types of information while allowing other kinds of signals through. A modification of this model was to suggest a more gradual processing, where even the filtered information could be accessed if it was important enough. For instance, even

if you aren't paying attention in a noisy room and somebody in the room mentions your name, you may hear it because this information is semantically salient to you.

Many other issues are involved in the cocktail party effect, including a principle called *auditory scene analysis* (in which acoustic events that are similar in frequency, intensity, and sound quality follow the same temporal trajectory in terms of frequency intensity, position, and so on). This principle may also be applied to speech. For instance, in a noisy room if you hear the words on a particular topic being uttered, say the weather, other words on this same topic may be more easily detected than random words relating to something entirely different. This is because when people talk about a certain topic, the listener often knows what will come next. For instance, if I tell you . . . "the American flag is red, white, and __," your chances of hitting the last word, *blue*, are really high here.

Much remains to be done to understand the cocktail party effect. This research is important for many applications, including the development of hearing aids and multi-party teleconferencing systems.

Chapter 14

Confirming That You Just Said What I Thought You Said

In This Chapter

- Discussing what makes speech special
- Exploring perceptual and linguistic phonetics
- Relating speech perception to communication disorders

Speech finally ends up in the ear of the listener. If nobody can hear it, there's no point blabbering about this or that or in measuring different kinds of sound waveforms. In the end, the difference between speech and other kinds of sounds is that speech conveys language and human listeners interpret it for language-specific purposes. Therefore, phoneticians study how people listen to speech and how speech fits into the bigger system of language.

This chapter attempts to answer some important questions. Here I discuss whether people listen to speech in different ways than they listen to other sounds. In addition, I address what people do when they listen to speech under less-than-ideal conditions. This chapter also covers the topic of what drives speech changes in language — the production or the perceptual side of things (or both). I also provide you a chance to apply this knowledge to the fields of child language acquisition and speech language pathology by considering how family members or other listeners may interpret (rightly or wrongly) the speech of children and brain-damaged adults.

Staging Speech Perception Processes

Researchers have proposed many different theories of speech perception over the years, and many will continue to develop. Perceiving speech begins with basic *audition* (hearing). Speech sounds are then further processed for acoustic cues, such as *voice onset time (VOT),* an important voicing feature of stop consonants in syllable-initial position. Phonetic information is then used for higher-level language processes. Check out the nearby sidebar for a glance at a couple of popular theories.

Eyeing speech perception theories: Bottom up, top down, or both?

A useful distinction in phonetics and other branches of the cognitive sciences is between information that is *bottom up* (information from the sensory/perceptual periphery) versus *top down* (based on prior experience/expectations). A bottom-up process involves information from the environment reaching you via the senses, and working its way up to your central nervous system. For instance, a bell rings and the sound waves hit your ear, causing electrical impulses to travel up the auditory nerve toward the brain. Eventually, the impulses reach the auditory regions of the cortex and you sense the sound.

A purely top-down process would be you just thinking (or imagining or dreaming) about a bell ringing. The same auditory region of your brain would likely be activated, but your auditory system wouldn't, and of course no bell would have to ring just because you imagined it. Top down refers to expectations and real-world knowledge and experience.

As you may have figured, top-down processing meets bottom-up processing on a daily basis. When you listen to speech, your mind generates expectations and predictions for what kinds of sounds will trigger phonetically meaningful units. The instant such sounds hit your ear means they're interpreted as speech. When you're listening to someone talk, your mind is constantly making (top down) predictions about what will come next, which is particularly helpful in noisy speech. These facts suggest that top-down information is crucial in speech perception. Of course, if you're deprived of your hearing, auditory speech perception immediately terminates. The take-home message is that both top-down and bottom-up processes are involved in a complex, interwoven fashion.

As phoneticians have learned more about how people perceive speech, certain key issues that require more attention have stood out. Researchers have noticed these issues, for example, when they weren't able to get computers or robots to do what humans can easily and effortless do. The following sections explain these special issues in speech perception.

Fixing the "lack of invariance"

This double-negative term, *lack of invariance,* simply means that the speech signal typically contains lots of variation, and yet human listeners are able to easily extract meaning from it. Put another way, there is a lack of one-to-one relationship between characteristics that scientists measure in the speech signal and the sounds that listeners perceive. Phoneticians know that listeners don't have the problem; scientists have the problem trying to figure out how people do it.

For example, most phoneticians agree that the formant frequency values are important cues for vowel quality. Chapter 12 lists the typical formant frequency values of /u/ for an adult American male (F1=353 F2=1,373 F3=2,321Hz).

F1 stands for first formant frequency, F2 for second formant frequency, and so on. However, it turns out that F2 for /u/ is higher when it follows an alveolar consonant, such as [t]. This effect is referred to as *coarticulatory* (also referred to as a *context dependent effect*). A coarticulatory effect occurs when the properties of one sound are influenced by the properties of an adjacent sound. In this case, the tongue shape for the back vowel /u/ is more fronted when the flanking consonant is an alveolar consonant /t/. This results in a higher F2 (second formant frequency) value. (Refer to Chapter 12 for more information on the relation between tongue position and formant frequencies.)

Figure 14-1 displays this effect, with /u/ and /tu/ side by side, left to right. The broad dark bands in the spectrogram (bottom half of the page) are the formant frequency estimates, with their midpoints shown by thin squiggly lines. If a phonetician were pinning her hopes on an invariant cue for /u/ in a defined frequency region of F2 space, she would get the sound dead wrong. That is, the second formant (marked by F2 in Figure 14-1) clearly starts higher in the /tu/ on the right side and has a different vowel formant frequency than in the case of the /u/ on the left. Something else must be going on. This example demonstrates a lack of invariance.

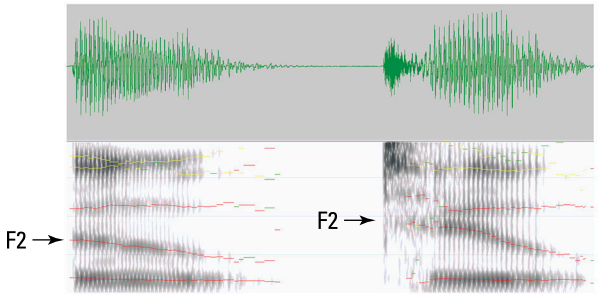


Figure 14-1: Comparing /u/ and /tu/.

Illustration by Wiley, Composition Services Graphics

Sizing up other changes

Another case of a lack of acoustic invariance in speech perception (which is so obvious that it sometimes escapes detection) is how listeners can understand the same thing said by many different people. I sometimes like to walk around my phonetics class and record ten different students saying the simple greeting "Hey!" When I later post the different spectrograms (refer to Chapter 13 for more on spectrograms), the dissimilarities between talkers are striking. Because of different vocal tract sizes, men and women differ. Also, the patterns of the [h] aspiration and vowel formant frequencies for [eɪ] can look quite different. The signals may have much variation, but anyone in the class can easily and effortlessly understand every single "Hey."

Taking Some Cues from Acoustics

A *cue* means information that a perceiver can extract from a signal. A speech cue is useful acoustic information taken from the spoken stream that a listener uses to interpret meaningful units of language (phonemes, syllables, words, and so forth). Phoneticians study how acoustic information may serve as cues for various sorts of meaningful categories. Chapter 12 covers some of the well-known acoustic cues, including formant frequency values for vowels and formant frequency transitions for consonants. Meanwhile, these sections introduce two important acoustic cues for consonants (VOT and burst characteristics) to show how listeners trade off when attending to different types of information that serve to designate similar phonetic categories.

Timing the onset of voicing

One significant cue to voicing in stop consonants is voice onset time (VOT). Listeners use VOT to tell whether a stop consonant is voiced at the beginning of a syllable, such as "*p*at," versus "*b*at," "*t*ad" versus "*d*ad," and "*c*oat" versus "*g*oat." VOT is a measure of time (in milliseconds) that elapses between the beginning of a *stop consonant* (the burst) and the onset of voicing. Long intervals of VOT correspond with stop consonants that sound voiceless, whereas short intervals sound voiced.

Figure 14-2 shows waveform examples for /dɑ/ (upper panel) and /tɑ/ (lower panel). You can see for voiceless /tɑ/ a relatively long lag (about 78 milliseconds) between the release of the "t" and the beginning of the vowel /ɑ/. For /dɑ/, the two events take place almost at the same time, about 11 milliseconds apart.

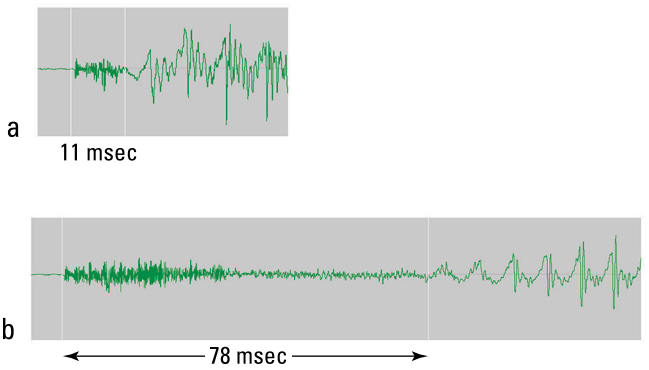


Figure 14-2: The VOT of /dɑ/ (a) and /tɑ/ (b).

Illustration by Wiley, Composition Services Graphics

Try your own VOT experiment and follow these steps:

1. **Place one hand in front of your mouth and under your lips (to feel aspiration), and the other hand above your Adam's apple (to feel your larynx buzzing) to get the sense of VOT under extreme conditions.**
2. **Make an *insanely* long voiceless "t."**

   Say "ttttttttttttttaaaaaa" as slowly as you can. Be sure to really sock the pronunciation of the "t."
3. **In between the blast of air for the initial "t" and the buzzing for the** /ɑ/, **let almost a half a second go by.**

   Include a lot of hissing air going out.

   Congratulations, you have made a 500-millisecond VOT.
4. **Say a regular** /dɑ/**.**

   Here, you should feel no hissing air, but should be able to sense the burst and buzzing taking place almost simultaneously.

In real life, English long-lag (voiceless) VOTs for syllable-initial consonants typically range from 40 to 100 milliseconds, with the averages increasing slightly as you move from labial (approximately 60 milliseconds) to alveolar (approximately 70 milliseconds) to velar (approximately 80 milliseconds) places of articulation.

You make stop sounds (/p/, /t/, /k/, /b/, /d/, and /g/) all day long. You somehow know that the initial voiceless stops will have long VOTs, and the voiced ones will be short. Precisely timed VOT values are important cues to let listeners know which stops you're intending (at least at the start of your syllables). You acquired these VOT values in childhood and have stuck with them ever since. Chapter 19 discusses what happens when VOT timing breaks down in the speech of people with communication disorders.

Bursting with excitement

Another unmistakable contender for an acoustic cue is the burst, the result of the release of air pressure for stop consonants. *Bursts* are very short events (about 5 milliseconds) that typically begin a stop consonant in syllable-initial position. Played by themselves, they pretty much sound like a stick hitting a table. However, bursts appear to have a lot of information packed into them.

Bursts are typically followed by a brief frication interval (approximately 10 to 20 milliseconds), as you can see at the far left side of the /tɑ/ waveform of Figure 14-2. Research has shown that stop bursts have unique spectral signatures revealing their place of articulation, which makes sense because the resonator in front of the source shapes the spectra. Such shapes would be quite different in the cases of, say, a /pɑ/, /tɑ/, and /kɑ/.

Experiments have shown that people and computers can use the information in stop consonant bursts to classify place of articulation with 85 to 95 percent success. Although researchers debate the theoretical importance of this finding, it's clear that listeners use such information to help determine the clarity of stop consonants.

To make your own burst (and to know that you're making one), record your speech online to check on your stop consonant production patterns. Use a program such as WaveSurfer or Praat to record yourself. Say "a*b*out" carefully, several times. Look at your productions in the raw waveform display. Does it show bursts for the "b"? Next, say "about" several more times, quite casually and loosely. Look to see what happens and try to willfully make your bursts come and go. Try to see if you're still understandable.

Being redundant and trading

A common letdown for beginning phonetics students is to notice that stops are frequently made *without* bursts. These so-called "burst-less wonders" occur more commonly in casual speech. Your challenge is to figure out how you, the listener, still know what you're hearing.

The answer lies in cue redundancy of speech production, and in the ability of listeners to engage in cue trading. *Cue redundancy* means that speech features are usually encoded by more than one cue, whereas *cue trading* refers to a listener's ability to hear more than one information source and sort out which cue is more important under different circumstances. A given phonetic feature, such as the voicing of a stop or the quality of a vowel, is rarely denoted by a single acoustic attribute. Instead, two or more sound attributes typically map onto a single feature, which would imply that if one cue (such as a burst) were missing, other acoustic cues could perceptually make up for it.

For instance, Chapter 12 notes how vowel quality (such as why /u/ sounds different than /æ/) is strongly conveyed by formant frequency values. However, other attributes can also play a role. For example, /u/ is generally shorter than /æ/ and is produced with higher pitch. The vowel /u/ also tends to have an off-glide quality, whereas /æ/ doesn't. These details illustrate cue redundancy: More than one type of acoustic information distinguishes /u/ from /æ/.

Under ordinary listening circumstances, some of these secondary factors may not weigh in as much as formant frequency values. However, if something masks or obscures a more usual cue, you may shift strategy and attend to some of the other data around. Welcome to the world of cue trading. Listeners engage in cue trading during speech perception, indicating listener flexibility. Figure 14-3 shows an example of cue trading in action.

Here the picture gets even more interesting with other types of sounds. Phoneticians have conducted a series of synthetic speech experiments

about what listeners tune in to exactly when listening to the difference between words such as "say" and "s*t*ay" (refer to Figure 14-3). Researchers created stimuli that signaled the "t" in the stop cluster "st" by the length of the silence (called a *stop gap*) in the cluster as well as a certain starting frequency of the F1 after the closure. When less of one cue is given to listeners, more of the other cue is required to give the same direction of response. For this "say/stay" example, when the stop gap is lengthened, leading listeners to a more "stay" response, an F1 can be higher. However, if the stop gap is shortened, the F1 must be lower for the same response. This response shows cue trading in action.

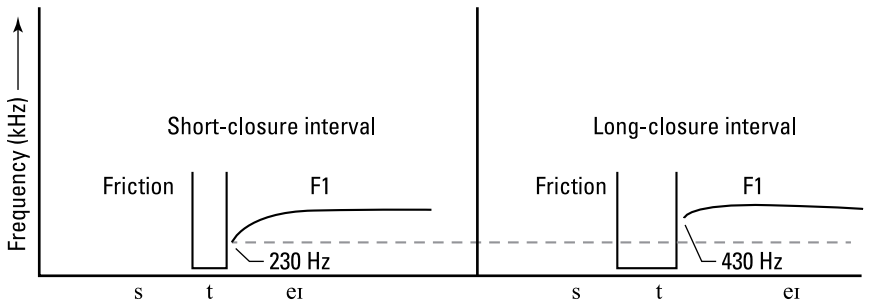


Figure 14-3: Cue trading in action.

Illustration by Wiley, Composition Services Graphics

Categorizing Perception

Perception refers to a person's ability to become aware of something through the senses (vision, smell, hearing, taste, and touch). Perception is different than *conception*, which refers to forming or understanding ideas, abstractions, or symbols. Perception is a sensory thing, while concept formation is a more mental thing.

In speech, you must perceive sound hitting your ear and rapidly interpret it so that you can use it for language. In one way, hearing speech is like hearing any other sounds (dogs barking, doors slamming, and such) in that it starts with your ear and goes to your brain. However, because speech is tied to language and communication, it seems to have some special properties. When you hear a speech sound, your brain doesn't have the luxury of sitting around and figuring out whether it's speech or not. Instead, your brain quickly makes a decision.

A type of behavior that has been widely studied in this regard is *categorical perception*, an all or nothing way of perceiving stimuli which actually vary gradually. The following sections examine categorical perception and show you how this special type of perceiving differs from other types of everyday perception. I also give examples of how categorical perception affects specific types of sounds and can play an important role in the classroom and clinic.

Setting boundaries with graded perception

Most perception isn't categorical. *Graded perception* is the typical type of perceiving you do when you sense something along a continuum. For instance, if someone gradually increases the intensity of the light in your bedroom (by turning up a dimmer switch), the room will gradually seem brighter to you. A graph of the intensity plotted against your reported brightness judgments should look like a happy upwards arrow, or more technically referred to as a *monotonic linear relationship* (refer to Figure 14-4).

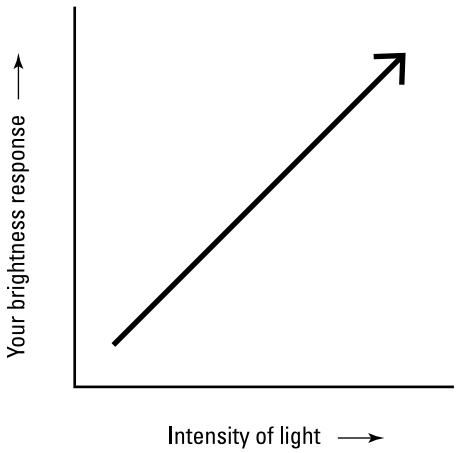


Illustration by Wiley, Composition Services Graphics

Figure 14-4: An example of a monotonic linear relationship.

This figure plots your brightness response on the vertical axis and light intensity on the horizontal axis. The greater the light intensity, the more you will report the light as seeming bright. This shows a hypothetical one-to-one ratio (monotonic) relationship between the physical (light intensity) and the psychological (how bright you say something is).

Now imagine you have a rather special friend in the room. Because he has spent many years as the stage director for a thrash metal band, something funny happened to his visual system and he now categorically perceives light. This is how your (fictional) friend would report the same event:

> "Dude! It's dark, dark, dark, dark . . . "
>
> "Now, I don't know . . ."
>
> "Okay, now it's bright, bright, bright, bright . . ."

Your friend doesn't respond with the (usual) graded series of judgments. Instead, he reports the following:

- A first series of intensities as "dark"
- A crossover point where he is basically lost (50 percent accuracy mean's he's unsure)
- A second series of intensities as "bright"

In categorical perception, even though stimuli are being adjusted gradually (such as by a dimmer switch) to the perceiver, it's as if the world is in one category or the other. A sharp flip occurs from one category to the next, and within each category the perceiver can't tell one stimulus from the next. Figure 14-5 shows a graph of this kind of function.

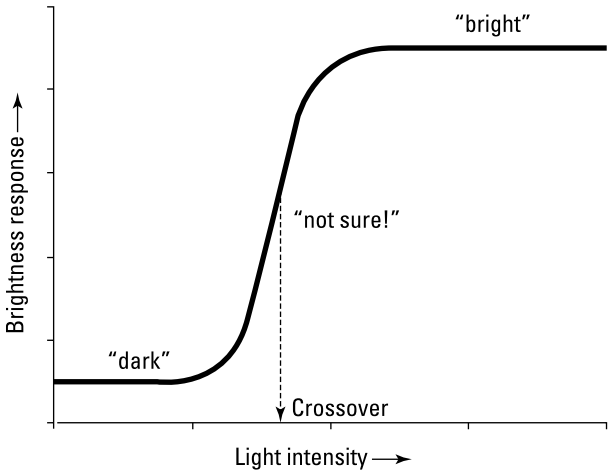


Illustration by Wiley, Composition Services Graphics

Figure 14-5: Sigmoidal (S-shaped) function for your strange friend's perception.

Instead of a linear monotonic relationship between graded stimulus and response, an S-shaped (sigmoid) function occurs. Start on the dark side, cross over to the light.

Here's how this example works for speech. In classic experiments conducted at Haskin's Laboratories in New Haven, Connecticut, researchers created synthetic speech as early as the 1950s by literally painting formants onto celluloid sheets that could be played back in a huge, scary device called the *pattern playback machine*. Using this kind of technology, researchers created synthetic speech stimuli, having a consonant burst and effective vowel onset that began at a specified point later. They then were able to create a continuum, beginning with VOTs increasing in equal steps from 0 to 60 milliseconds. Very short-lag stimuli should sound maximally like /dɑ/ and long-lag stimuli should sound most like /tɑ/. Figure 14-6 shows what these stimuli might look like.

If I played the stimuli shown in Figure 14-6 to you in equal steps and you heard things in a graded fashion (like say, dog barks or ringing bells), then you would expect between each step the same amount of change in /dɑ/ to

/tɑ/ judgment, giving rise to a linear function if one were to plot your hearing against the stimuli themselves.

However, that's not what occurs with VOT identification. Instead, listeners report stimuli having VOTs of 0, 10, 20, or 30 milliseconds all being 100 percent good /dɑ/. If a stimulus is played that is about 35 milliseconds long, listeners are confused, calling half of them /dɑ/ and half of them /tɑ/. By about 40 milliseconds, most stimuli are called /tɑ/. After about 40 milliseconds, everything is completely /tɑ/. It's as if there is a /dɑ/-land to the left, a /tɑ/-land to the right, and a no-man's zone in between. Refer to Figure 14-7.

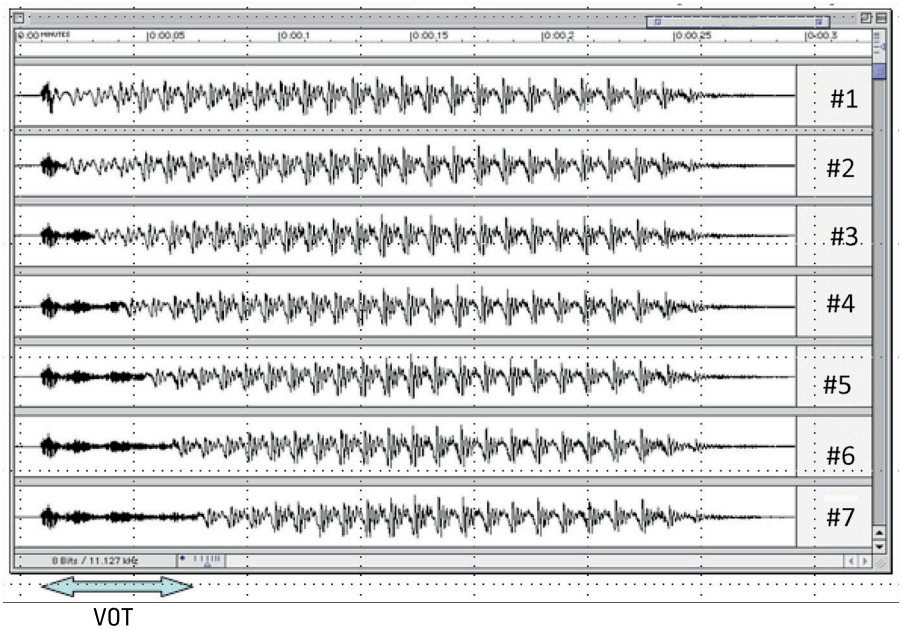


Figure 14-6: Sample synthetic speech stimuli used for VOT listening experiments.

Illustration by Wiley, Composition Services Graphics

Understanding (sound) discrimination

The flip side to this fascinating type of listening is when people are asked to *discriminate* (say "same or different") between stimulus pairs. *Sound discrimination* is a task in which the listener doesn't need to name or identify anything, but instead judges two or more items as same or different. People can usually discriminate many more different sounds than they can identify. Figure 14-8 shows a graph of the data. Take a look at the far left side of the graph: When listeners must say "same" or "different" to two stimuli with either 0 to 10 or 10 to

20 msec combinations, they perform poorly. That is, they can't tell any of these pairs apart (both members will likely sound like perfectly good /dɑ/).

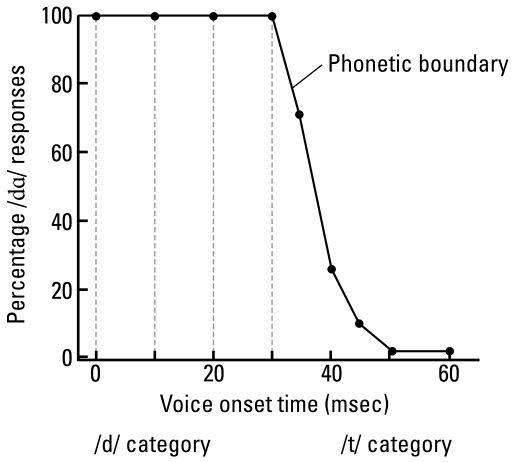


Illustration by Wiley, Composition Services Graphics

Figure 14-7: Plotting /dɑ/ and /tɑ/ identification.

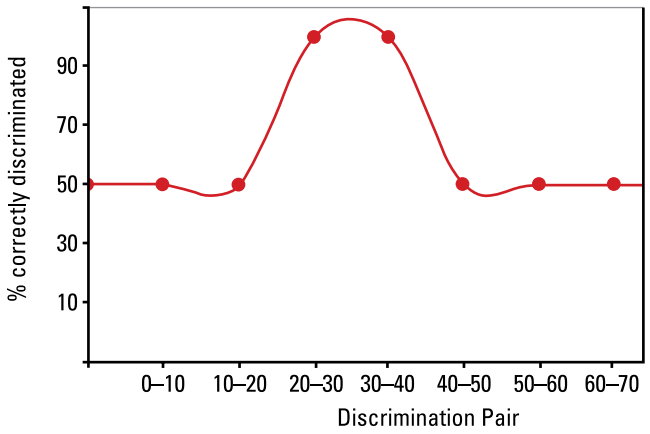


Illustration by Wiley, Composition Services Graphics

Figure 14-8: Plotting /dɑ/ versus /tɑ/ discrimination.

They are in /dɑ/-land. At the far right of the graph, you can see the same pattern: Listeners can't tell the difference between any of the good /tɑ/s. The listeners are in /tɑ/-land; they all sound the same to them. There is no such thing as a good /tɑ/ or a bad /tɑ/. However, in the middle of the graph, you can see what takes place when one member of the pair falls within the short-lag boundary (/dɑ/-land) and the other on the long-lag boundary (/tɑ/-land). Here, listeners can distinguish quite well between the pair, with discrimination at almost 100 percent.

Examining characteristics of categorical perception

Categorical perception applies to many cues in speech. VOT is just one example. Table 14-1 shows some other examples.

Table 14-1 Examples of Categorical Perception

Feature	*Cue*	*Example*
Final consonant voicing	Duration of preceding vowel — longer before voiced final consonant	/bæt/ versus /bæd/
Place of articulation — oral stops	Start and direction of F2: **Bilabial:** Starts low in frequency and goes up to vowel F2 value. **Alveolar:** Starts around 1800 Hz and goes to vowel F2 value. **Velar:** Starts high in frequency, goes down to vowel F2 value.	/bɑ/, /dɑ/, /gɑ/
Place of articulation, nasal stops	Start and direction of F1 and F2	/mɑk/ versus /nɑk/
Voicing in final fricatives	Duration of preceding vowel — longer before voiced final consonant	/ɑs/ versus /ɑz/
Place in fricatives	Frequency of noise hissiness — higher in /s/ than /ʃ/	/sɑ/ versus /ʃɑ/
Liquids	Frequency of F3 — lower before /ɹ/ than /l/	/ɑɹ/ versus /ɑl/

To get a sense of how people categorically perceive different sound contrasts, begin by looking down the Feature column on the left in Table 14-1. The Cue column shows the attribute that categorically varies. An example (in IPA) is provided on the far right.

For instance, glance down to the second entry in the Feature column. The Cue information notes that listeners categorically hear differences in the start and direction of the second formant frequencies (F2). Refer to Chapter 12 for more information on formant frequencies as cues to consonant place of articulation.

Categorical perception is crucial to the fields of phonetics and psycholinguistics. Here are some important things to keep in mind about this intriguing aspect of our human behavior:

- When researchers first uncovered these effects in synthetic speech experiments in the 1950s, they thought categorical perception was unique to humans.
- Categorical perception has since been demonstrated in the communication systems of bullfrogs, chinchillas, monkeys, bats, and birds.
- Some auditory theorists take issue with some of the categorical perception experimental findings and instead suggest that more general auditory (non-speech) explanations may account for the results. They reject the idea of a special module for speech perception.

The following are some ways categorical perception plays a part in phonetics.

How people master second languages

Categorical perception is language-dependent and therefore experience-based. *Monolingual* speakers (people who only speak one language) acquire these boundaries at an early age (typically 9 to 12 months old). Children raised bilingually map the acoustic patterns of the languages they acquire in a separate fashion and are able to keep them reasonably distinct (more research needs to be done in this area). Adults learning a second language face an interesting dilemma: They must overcome the perceptual boundaries of their native language (L1) in both perception and production, in order to become proficient users of their second language (L2).

This raises an interesting issue: Have older L2 language learners missed out on something with respect to language learning? That is, because phonetic categories are important to how people learn language and these categories are formed early in life, are older second-language learners in a difficult situation with respect to language learning? And is good accent acquisition age-dependent? Evidence supporting this depressing idea seems to be everywhere, such as the immigrant family that has just arrived where Grandpa can't speak English at all, but little Junior already sounds like he was born in his new country. Also, empirical studies on the relationship between age and accent generally support this view.

Many factors clearly influence who becomes successful in second language acquisition and why, including cultural, social, and motivational. Other factors may include an inborn propensity or talent for speech and language learning, and an age factor, called a *critical* (or *sensitive*) *period.* Although a critical period doesn't seem to be the case necessarily for the acquisition of syntax, vocabulary, and other more mental properties of language, it just may be the case for native-sounding accent.

In speech and language pathology

Another important application of studying categorically perceived phenomena (such as VOT) is to the world of speech and language pathology. Chapter 19

describes the main symptoms of Broca's and Wernicke's type aphasia. Broca's aphasia results from left anterior brain damage and leaves patients with poor speech output and generally good comprehension. Wernicke's aphasia is marked with fluent, semantically empty speech, and poor comprehension.

Studying the VOT of stop consonants produced by these subjects has provided important information about the nature of their problems. Although both types of aphasic individuals each make speech sound errors (for example, saying "Ben" instead of "pen"), scientists now assume that the errors of Broca's aphasia subjects come largely from problems with mistiming and coordination problems, whereas the Wernicke's aphasia patients substitute incorrect (but well formed) sounds. Take a look at their VOT patterns in Figure 14-9 to see why.

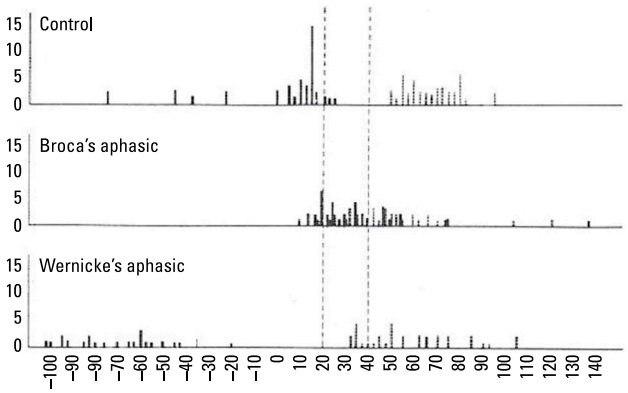


Figure 14-9: VOT values for three adult talkers: Control, Broca's aphasic, and Wernicke's aphasic.

S.B. Filskov and T.J. Boll (Eds), Handbook of Clinical Neuropsychology, J. Wiley & Sons, 1981. This material is reproduced with permission of John Wiley & Sons, Inc.

In this figure, called a *histogram* (a bar chart that shows frequencies), VOT values are plotted for /dɑ/ and /tɑ/ syllables made by a healthy talker (top), an individual with Broca's aphasic (middle), and a person with Wernicke's aphasia (bottom). The healthy adult shows a cluster of /dɑ/ values centering around 10 milliseconds (arrow), with a few pre-voiced instances farther to the left. Meanwhile, on the /tɑ/-side, long-lag VOT center around 65 milliseconds, with some productions going as high as 90 milliseconds. Therefore, the healthy talker has two different sets of stops, those with long lags and those with short lags.

By contrast, the person with Broca's aphasia seems to be in trouble. His VOTs don't fall into the two usual categories, but instead fall into the no-man's land (marked by dotted lines) in which most listeners can't hear the difference between a /dɑ/ and a /tɑ/. You can predict that the mistiming of these aphasic talkers can get them into big perceptual trouble when other listeners hear their speech (refer to Chapter 12 for more information).

The productions of the Wernicke's aphasic talker, like those of healthy adults, show /dɑ/ and /tɑ/ VOT values in two distinct categories. This suggests that any errors coming from them are likely substitutions, not mistimings.

Balancing Phonetic Forces

Phoneticians must be able to explain why talkers may sound different in various (such as formal versus relaxed) speaking situations, but why these kinds of speaking adjustments don't change people's speech so much that people become less understood. Phoneticians must also explain how a language may change its sound system over time. In this section, I discuss two principles designed to address these issues: Ease of articulation and perceptual distinctiveness.

Examining ease of articulation

Ease of articulation is the principle by which speakers tend to use less physical effort to produce speech. This, in turn, can affect sound change in words. English pronunciation has many examples. Consider, for example, how "often" is usually pronounced without a "t." Such a sound drop is called an *ellipsis*, where a part of a consonant cluster is eliminated. Chapter 2 mentions how speaking involves a balance between getting your words out in time and with the least effort, on the one hand, and making yourself understood, on the other.

Over time, you may expect that such pronunciation changes could cause the spelling for a word to eventually switch. You have already noted an example of this kind of thing happening with the word "impossible," in which the prefix "in" changed to an "im" to allow *assimilation*, the sharing of features that are easier to say together. In this case, the shared feature is the bilabial place of articulation of the /m/ and /p/ consonants. This change actually occurred fairly early in the history of English.

Ease of articulation is a concept that is simple to grasp, but tricky for experts to precisely define. In general, vowels are easier than consonants. Also, the fact that infant babbling begins with consonant-vowel (CV) syllables suggests that certain syllable types are more basic than others.

Another interesting source of information is *diachronic* (across time) evidence, describing how languages change in history. For example, modern Spanish, just like English, doesn't have phonemic vowel length. Therefore, the word "casa" is no different than "caaaaaasa," they both mean house. A

longer vowel is more difficult to produce than a shorter vowel because of the extra time and energy spent to expel air out of the lungs. Thus, ease of articulation played a role in changing the vowel system of Spanish.

Ease of articulation also applies to sign languages, indicating that such processes are more general than sound-based articulatory systems. For instance, the study of American Sign Language (ASL) and German Sign Language (Deutsche Gebardensprache, DGS) has shown that the most fluent signers tend to make more *proximal* (closer to the body) movements in order to maximize skill and comfort. This may suggest that over time more *distal* (away from body) gestures would be moved closer to the body.

Focusing on perceptual distinctiveness

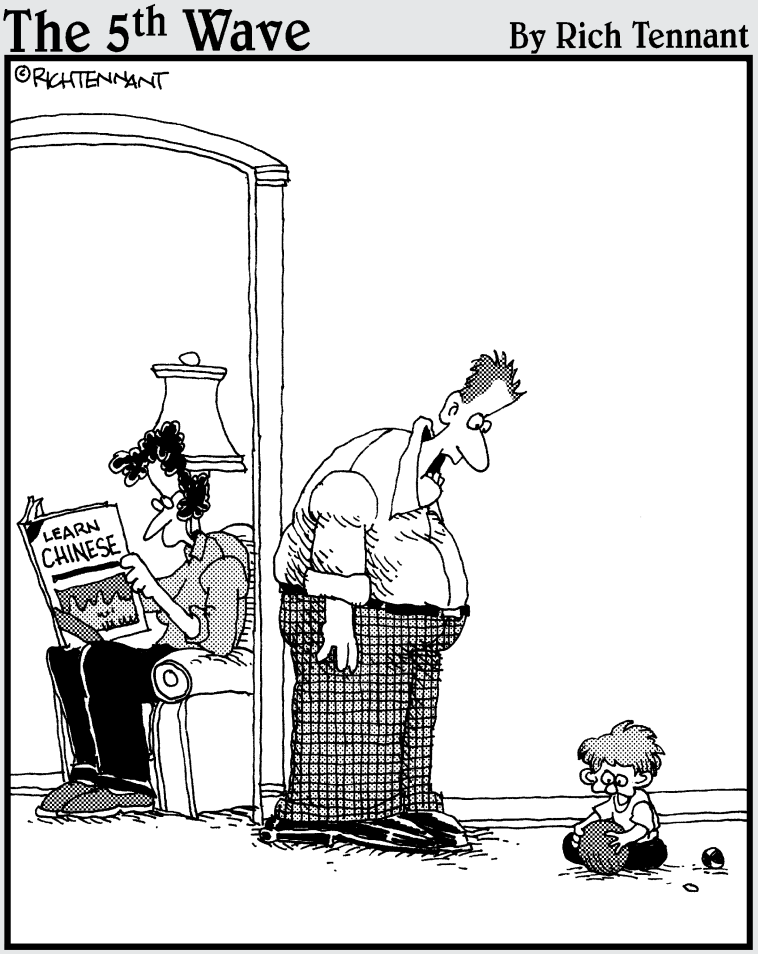
People can't be lazy with their articulators forever and get away with it. Other people are listening, which explains why being perceptually distinct is important. *Perceptual distinctiveness* is a property critical to language because languages can't have words so close together in sounds that people can't tell them apart. To be sure that such a confusing situation doesn't take place, a language must ensure *sufficient perceptual separation,* which in layman's terms means the sounds of a language are different enough that they can be heard as such by listeners. If a language has a certain sound in its inventory, then the nearby sounds must be distinct; otherwise pandemonium can result. Perhaps the easiest way to see the importance of this property is to take a peek at the vowel systems of the world's languages.

Linguists have sampled major language families (and subfamilies) of the world's languages. One of the most extensive databases is the UCLA Phonetic Segmental Inventory Database (UPSID), a collection of more than 317 languages. From a survey of the world's verbs, linguists have discovered the following distinctions about the world's languages' vowel systems:

- Languages seem to use anywhere from 3 to 15 vowel phonemes in their inventory.
- Five-vowel systems (such as Latin, Spanish, Japanese, Swahili, and Russian) are the most common. For these vowels, the typical inventory is /i/, /e/, /a/, /o/, and /u/.
- Vowels tend to distribute in symmetrical ways and fill out the space of the vowel quadrilateral. Thus, no five-vowel language consists of only closely grouped, front vowels, such as /i/, /ɨ/, /ɪ/, /e/, and /ɛ/.
- Distinctions such as length (short versus long) and nasalization (as in French) are more common in languages with a large number of vowels than with small vowel inventories. This theory suggests that such features can help keep things clear in a more crowded vowel space.

Part IV

Getting Global with Phonetics



"You mean, 'wo,' 'ta,' 'baba,' and 'mama' are all words in Mandarin? My gosh, Alice, our baby's been speaking Chinese the last few weeks!"

web extras

Visit dummies.com/extras/phonetics for more great Dummies content online.

In this part . . .

- Understand how the world's languages can differ by airstream mechanisms, voice quality, and tone.
- Grasp how different languages use different manners of articulation, including glottal, trills, and taps, and what you need to know in order to produce these sounds yourself.
- Differentiate between a *dialect* and *accents* so you can identify different varieties of the same language.
- Identify a wide array of English accents, from the various American and English accents to Canadian, South African, Australian, New Zealand, and more, to help you distinguish one variety from another and grasp how they involve different sounds.
- Examine when children and adults have speech and communication issues and when speech errors may require professional help.

Chapter 15

Exploring Different Speech Sources

In This Chapter

- Getting familiar with language families
- Experiencing airstream mechanisms
- Tuning up your ears to tone
- Detecting new voice onset time (VOT) boundaries

All speech starts on a breath stream. To fully appreciate the amazing variety of ways that people can make speech sounds, it's important to look (and listen) beyond English. This chapter begins with a discussion of the different types of airstream mechanisms people use to produce speech. I next take you on a tour of phonemic tone, a sound property foreign to English but quite common in the languages of the world. The chapter wraps up with an introduction to some very different states of the glottis for speech, including breathy voice and creaky voice.

Each new language sample is paired with links to online audio and practice exercises. These samples give you hearing and speaking experience, in order to make this more real.

Figuring Out Language Families

This chapter (and Chapters 16 and 17) introduces you to some sounds in other languages of the world. For this information, it's helpful to know how linguists group languages. A *language family* is a group of languages that descend from a common ancestor. If you can work with the idea of a family tree, you can easily work with a language tree.

Figure 15-1 gives an example for English. At the base of the tree is Proto-Indo-European, a hypothesized proto-language thought to be the precursor of many languages found today in Europe and the Indian subcontinent.

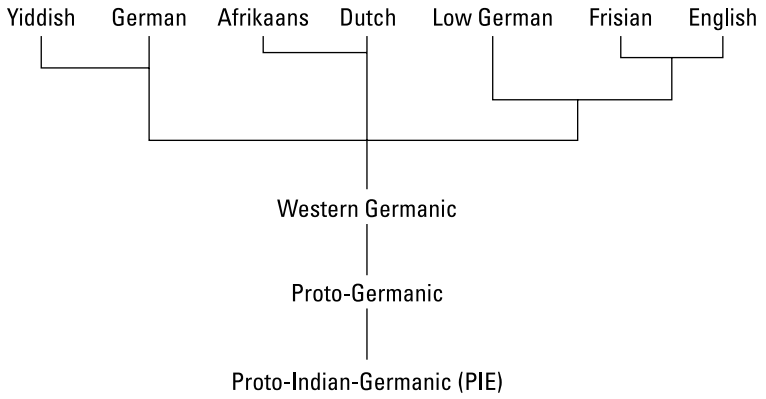


Illustration by Wiley, Composition Services Graphics

Figure 15-1: A language family tree for English.

Nobody really knows who spoke Proto-Indo-European (PIE) or exactly when. One theory projects potential speakers of PIE somewhere between 8000 and 4000 BCE. They may have lived near the Black Sea in Russia or in Anatolia (modern day Turkey).

Moving up the tree, you arrive at the Proto-Germanic branch. The speakers of this proto-language were thought to live between 500 BCE to 200 CE, in regions comprising southern Sweden and modern-day Denmark. Climbing up the tree from there, you reach the Western German branch. At this point, branches split into English, Frisian, Low German (Saxon), Dutch, Afrikaans, German, and Yiddish. Technically, West-Germanic is a mother language of English, while its sister languages are Frisian, Low German (Saxon), Dutch, Afrikaans, German, and Yiddish.

According to the Dallas-based Summer Institute of Linguistics (SIL), there are approximately 6,900 world languages. Recent estimates suggest about 250 established language families can be used to group these languages. The good news is, nearly two-thirds of these languages (accounting for ⅔ of the world's population) can be accounted for in a top six grouping of families. These groupings are as follows:

- **Niger-Congo:** Approximately 350 million speakers, accounting for 22 percent of the world's languages. Most widely spoken are Yoruba, Zulu, and Swahili.
- **Austronesian:** About 350 million speakers, accounting for 18 percent of the world's languages. Most common are Tagalog, Indonesian, and Cebuano.

- **Trans-New Guinea:** Three million people speak 7 percent of the world's languages, including Melpa, Enga, and Western Dani.
- **Indo-European:** Three billion people speak 6 percent of the world's languages, including English, Spanish, Hindi, and Portuguese.
- **Sino-Tibetan:** About 1.2 billion people speak 6 percent of the world's languages, including Mandarin, Cantonese, and Shanghainese.
- **Afro-Asiatic:** Approximately 350 million speak 5 percent of the world's languages, such as Arabic, Berber, and Amharic.

These language families are the largest because of the number of languages in each family. This doesn't mean the largest number of speakers speaks them nor does it mean they have the largest geographic spread.

Eyeing the World's Airstreams

An *airstream mechanism* is how air is set in motion for speaking. In this section, I ground you in the physiology of English speech by describing how consonants are produced by air flowing outward from the lungs. I also look at more unusual mechanisms (from the throat and the mouth) that can result in very different sound qualities than are typically used in English. Airflow will in some cases be directed into your body. However, please don't worry. I promise it will be fun, legal, and nobody will get hurt.

Your master guide to this next section is Figure 15-2. This figure summarizes the airstream mechanism by airflow direction and anatomy. You can use this figure to identify some of the different sounds of the world's languages based on which airflow direction and part of the vocal tract are used.

	Pulmonic	Glottalic	Velaric
Egressive	Plosives /p/, /t/, /k/, /b/, /d/, /g/	Ejectives /p'/, /t'/, /k'/	None
Ingressive	None	Implosives /ɓ/, /ɗ/, /ɠ/	Clicks /ʘ/, /ǀ/, /ǃ/, /ǂ/, /ǁ/

Figure 15-2: The airstream mechanisms by airflow direction and anatomy.

Illustration by Wiley, Composition Services Graphics

Going pulmonic: Lung business as usual

An *egressive* (outward) airflow that is *pulmonic* (from the lungs) is the most common airstream mechanism. Even languages with airflows that are temporarily made in other ways default to outgoing lung airflow, most of the time.

Take a moment and think about your breathing. Inhalation begins when you actively contract muscles, notably the *diaphragm* (a large, dome-shaped muscle at the base of the lungs). This causes the chest cavity to get bigger, and air to rush into your lungs. Although you don't usually need to consciously think in order to start this process, energy is certainly required.

In contrast, during exhalation you're usually letting go, which is a passive process. During speech, people take sharper inhalations and hold back their exhalations in order to maintain a long and steady flow of air to speak on. If you imagine having to hold a long note while singing (or playing a woodwind, such as flute), you can get the idea of why speaking needs a long-lasting, outgoing airflow. The lungs supply that airflow.

Examples of stop consonants made on the pulmonic egressive airflow include the plosives /p/, /t/, /k/, /b/, /d/, and /g/ — all found in English. These consonants get their name from the fact that they're produced with an explosive quality when the articulators are separated, marked by a sudden release of air (not a long-lasting outflow of air).

Considering ingressives: Yes or no?

What about ingressive airflow, producing speech sound by sucking air in from the lungs? Possibly, but this method isn't used regularly for language. You can say it's used *paralinguistically*, meaning it's related to the nonverbal parts of language use. For instance, in Scandinavia, "ja" (*yeah*) ingressive sounds are used for conversational backchanneling. *Backchanneling,* like nodding, is letting your speech partner know you're paying attention (or at least, pretending to). However, no self-respecting Swede or Dane would say phrases or sentences on an inhaled pulmonic airstream.

A *pulmonic ingressive phoneme* was found in an Aboriginal ritual language, Damin. This magical language of Shamans had chants and incantations of every known breath mechanism. Such behavior exceeds that of even the most enthusiastic New Age devotee in Zurich or New York City. Unfortunately, the last speaker of Damin died in the 1990s. Somehow, there doesn't seem to be a bright future in pulmonic ingressives.

Talking with Different Sources

If you're a native speaker of English, some foreign speech airflow mechanisms may be a bit outside of your comfort zone. You likely don't say a lot of things by pushing air back and forth from your glottis or by clicking around in your mouth. However, many millions of speakers in the world do. These sections identify three types of sounds created by non-English different airstream mechanisms: implosives and ejectives, and velarics.

Pushing and pulling with the glottis: Egressives and ingressives

The *glottalic* (produced by actions of the larynx at the glottis) airstream mechanism allows talkers to add emphasis to certain sounds by a piston-like action of the vocal tract. Here is how it works: In egressive stop consonants (also known as *ejectives*), the glottis clamps shut and pushes air up and out of the mouth like a bicycle pump using a cylinder action, which gives stop consonants a certain popping quality. Because the glottis is tightly closed, no air can escape to cause vibration, therefore all ejectives are voiceless.

In ingressive stop consonants (also known as *implosives*), the glottis closes and then moves down, pulling air into the vocal tract. The narrow opening in the glottis allows air to move upward through it, creating slight voicing. This is like the bicycle pump working in reverse. Implosives have a peculiar sound. Figure 15-3 shows the mechanics involved.

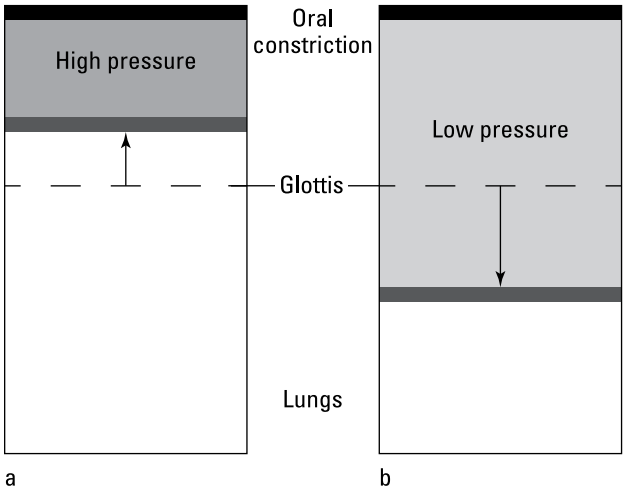


Figure 15-3: Mechanics of egressives (a) and ingressives (b).

Illustration by Wiley, Composition Services Graphics

In Figure 15-3a, the glottis is completely shut, which creates a high pressure. In Figure 15-3b, the glottis narrows for downward suction, while still slightly open for voicing, creating a lower pressure.

Overall, ejectives are more common than implosives. They're also easier to produce. Feel like making an ejective? A famous way of making one, suggested by the phonetician Peter Ladefoged, is to do the following:

1. **Hold your breath.**
2. **While still holding your breath, try to make a "k" as loudly as you can.**
3. **Relax and breathe again.**

   Congratulations! You have just made an ejective /k'/.

Velars, by the way, are the most common place of articulation for ejective sounds. Chapter 4 discusses velars in greater depth.

Some languages that have ejectives are Hausa (West Africa), Quechua (South and Central America), Lakhota (Sioux), Navajo, and Amharic (North Africa). Some languages that have implosives are Sindhi (Pakistan, India), Igbo (West Africa), and Paumari (Brazil).

Carl Sagan: Astronomer, educator, and poster boy for implosives

Carl Sagan (1934–1996) was an American astrophysicist, cosmologist, author, and science personality beloved for his role in educating the public about outer space. He was a professor of astronomy at Cornell University and later became popularly known for the 1980s television series *Cosmos: A Personal Voyage*. Dr. Sagan was also teased for his use of the word *billions*, as in *billions upon billions of stars*. Sagan purposely emphasized his pronunciation of *billions* to distinguish it from *millions*. His rather affected pronunciation led to frequent satires by comic performers. You can find a sample at `www.youtube.com/watch?v=1jVQg87MA9s`.

In his honor, his colleagues have suggested (humorously) that a measure called the *Sagan* should refer to *a large amount of anything.*

Phonetician Peter Ladefoged noted that Dr. Sagan produced an implosive bilabial at the beginning of the word. It turns out that Sagan's pronunciation of *billion* shared something with millions of Sindhi and Igbo speakers — briefly sucking in air by lowering the glottis while producing the initial stop consonant. Had others not beaten phoneticians to posthumously naming something a *Sagan*, the phoneme /ɓ/ could've been named in his honor!

To hear Carl Sagan say the "b" word, visit `www.utdallas.edu/~wkatz/PFD/carl_sagan_billions.mov`

Producing implosives can be a bit tricky. Most people don't feel comfortable voicing while breathing in. Let me suggest some steps to work on implosives:

1. **Take a deep breath and say "aah" while inhaling.**

   Your voice should sound scary, as if you're in a horror film.

2. **Now say "bah" in a regular manner.**

   Can you say "bah" while breathing in?

3. **Work on inflowing breathing for the "b" alone, while the rest of the sound is made with a regular outward air flow.**

   Congratulations! You have made (or at least started to make) an implosive bilabial stop, /ɓ/.

Some people do better by imitation. Check out these samples from Sindi at `www.phonetics.ucla.edu/course/chapter6/sindhi/sinhi.html`.

Here is a great suggestion for making the implosive /ɠ/: First make the "glug glug" sound for chugging down a drink. This typically lowers the larynx. You can then transfer this gesture to /ɠa ɠɑ/, /ɠu ɠu/, and other vowel contexts.

Clicking with velarics

The third airstream mechanism, *velaric* (a click produced from the velum) is certainly the most thrilling. You can find these clicks in many languages spoken in South Africa. To form a click, the speaker produces a pocket of air within the mouth and then releases the air inwards. Placing the tongue back against the velum creates a mini-vacuum, which is then released by the front of the tongue to cause clicks having different places of articulation in front of this velar closure. Figure 15-4 shows an example of the stages of producing an alveolar click. This sound is like the "tsk-tsk" (as in "shame on you!") noise created by placing the tongue behind the teeth.

Clicks have different places of articulation and their own special symbols in the IPA (although Roman letters such as "c," "q," and "x" are used for spelling clicks in African languages like Xhosa and Zulu).

These sections break down two types of velarics.

Making a bilabial click

A first click to try is the *bilabial*. Really, anyone can make this sound. This is a kind of "kissing sound," but remember, it's a consonant that is followed by a vowel. Conveniently, the IPA symbol is something that looks like a round mouth /ʘ/ kissing.

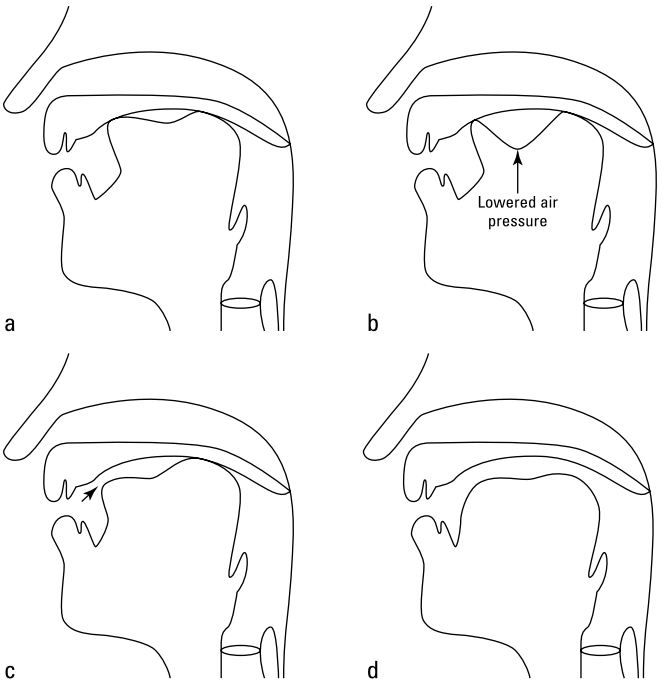


Illustration by Wiley, Composition Services Graphics

Figure 15-4: Producing alveolar clicks.

To make this bilabial click sound, try these steps:

1. **Put your forefinger to your lips and make a light kissing sound.**
2. **Try this sound followed by the vowel** /ɑ/**.**
3. **Now put it in medial context,** /ɑʘɑ/**.**
4. **Explore other vowel contexts such as** /uʘu/ **and** /iʘi/**.**

Bilabial clicks are quite rare among the world's languages. !Xóõ, a language spoken in Botswana, has this sound. You can find an example at `www.phonetics.ucla.edu/course/chapter1/clicks.html`.

Making a lateral click

Another non-linguistic click sound people commonly make is a lateral click /ǁ/, for encouraging a horse to hurry up. To hear a broad range of click sounds in speech, visit `www.phonetics.ucla.edu/course/chapter1/ipaSOUNDS/Con-58b.AIFF`. I also recommend listening to examples of click sounds produced in word contexts. You can find some Zulu at `www.youtube.com/watch?v=MXroTDm55C8`.

Just to show that a monolingual English speaker can pick up these sounds, here is a Texas student reciting a famous Zulu poem about a skunk and a tale about an Iguana: `www.utdallas.edu/~wkatz/PFD/skunk_iguana.wav`.

Putting Your Larynx in a State

Most people take for granted that they can speak with their larynx vibrating in the same basic way each time. Some voice coaches call this *chest register* — the vibratory patterns you use for everyday speaking. You might step outside of this state in order to sing high (for instance, falsetto), to whisper, or to try and project your voice down extra low (creaky). None of these laryngeal changes affect meaning in English. However, in some of the world's languages, the way in which you vibrate is the way you get your message out.

In this section, I identify two states of the glottis used to change meaning in a number of languages throughout the world.

Breathless in Seattle, breathy in Gujarat

Breathy voice (or *murmur*) is a state of the glottis in which the vocal folds are slightly more open than usual, as the result of high airflow. In breathy voice, the folds vibrate while they remain apart. The result is an "h"-like sound that has a kind of sighing quality. This breathy "h" sound is written in IPA as /ɦ/. It occurs in English in words such as "behold" or "ahead," although people don't hear it as such. In many languages in India, murmur plays an important (phonemic) role.

For instance, Gujurati, a language with approximately 66 million speakers, distinguishes plain and murmured sounds. The IPA symbol for murmured voice is two dots placed beneath the symbol [◌̤]. Stops can also be produced with a murmured release, indicated with a diacritic consisting of a small breathy h to the upper right [ʱ]. For example, [bʱa̤r] means "burden" and [bar] means "twelve."

Croaking and creaking

Creaky voice (also known as *laryngealized* or *vocal fry*) is a very low-pitched variation that has a rough, popping quality. In creaky voice, the vocal folds are positioned rather closely together except for a small top opening. This position allows the vocal folds to vibrate irregularly in a manner that produces a characteristic raspy sound when air passes through.

Creaking as the new cool?

Creaky voice is rather noticeable and can be introduced into singing styles to give a characteristic quality to the voice (think of Britney Spears, Ke$ha, and Lady Gaga). Other media personality figures such as the Kardashian sisters have a creaky quality.

Recent studies have led some linguists to question whether creaky voice may be picking up in the United States, perhaps as a fad among young, white, urban upscale women. Some surveys have shown a surprisingly high rate of fry register in the voices of these subjects (in one study, more than two-thirds of the 34 young American women sampled). Creaky voice was formerly thought to be more prevalent in men. For example, a British survey showed higher creaky voice usage among men. Some sociolinguists suggest that an emphasis on creaky voice may be a kind of group bonding thing, where participants indicate they're hanging out with the right crowds. At this writing, the sample sizes are small and it isn't certain exactly what's going on.

Many people naturally have creaky voice as their voice trails off. This sound quality can be increased by damaging your vocal folds (such as smoker's voice) or through conscious effort and practice, as in certain types of singing (pop, country western, gospel bass).

In English, saying "hello" (regular) or "hello" (creaky) would tell a listener nothing new, except perhaps your mood. However, a number of West African languages (including Hausa and Yoruba) use creaky voice to distinguish meaning. The IPA diacritic for creaky voice is a tilde placed under the sound, like this [̰]. For example, from the Mixtec family of languages in Southern Mexico, [kḭnin] means "tie down," whereas [kinin] means "push."

Toning It Up, Toning It Down

In phonetics, *tone* (also known as *phonemic tone*) refers to when the pitch of a sound changes meaning. This definition is a more specialized use of the word "tone" than when people make comments such as "I don't like the tone of his voice" (meaning the emotional quality conveyed). This specialized use of tone also doesn't refer to the melody of language over larger chunks of speech, such as the rising quality at the end of some questions in English. These broader aspects of language melody, known as *sentence level intonation,* are discussed further in Chapters 10 and 11.

If you're a native speaker of English (or most other Indo-European languages for that matter), you don't have phonemic tone. I hate to break it to you, but

linguistically you're the odd man out, because most of the world's languages are *tone languages* (languages having phonemic tone). If you fall in this non-tonal category, taking a look at how languages handle phonemic tone in these sections can be helpful.

Register tones

The simplest tone languages are called *register tones,* having relatively steady pitches and levels, such as high, medium, and low. The simplest cases are two-toned systems, high versus low (as in many Bantu languages, including Zulu). Many languages have three-way (high/mid/low) systems (for instance, Yoruba), although languages with four- and even five-way systems exist.

Register tone languages have a default (basic) tone, against which the other tones contrast. Languages that don't have phonemic tone (like English) are considered *zero tone* languages, with other kinds of pitch contrasts used instead.

The IPA has a few ways of indicating register tone (see Chapter 3 to refresh yourself on the IPA chart). The easiest system is to use diacritics placed over the vowel: An *acute* accent (slanting left) indicates high tone [ˊ], a level mark (macron) means mid tone [ˉ], and a *grave* accent (same direction as a backslash!) means low tone [ˋ].

For example, look at these three different tones from Akan Twi, a language spoken in two-thirds of Ghana, Africa:

[pápá] means "good" with high-high tone.

[pàpá] means "father" with low-high tone.

[pàpà] means "fan" with low-low tone.

Contour tones

In languages with *contour tones,* at least some of the tones have movement (or direction). Most typical is a simple rising or a falling pitch. Some movement patterns can be more elaborate, such as the dipping pattern in Thai or Mandarin.

A useful language to examine to get a handle on contour tone is Mandarin Chinese. In most standard dialects, Mandarin has a four-tone system, as shown in Figure 15-5. In contour tones, a speaker's goal is to produce pitch movements, rather like hitting a target. Unfortunately, the spelling system used to transcribe tones in Chinese speaking countries, called *Pinyin,* is in a very different order than the IPA. In this figure, I also describe the pattern so that it remains clear. ***Note:*** Changing the tones can make different words.

Tone Number	Chinese Character	IPA Tone Symbol	Tone Description	English Translation
1	媽	˥	High level	“mother”
2	麻	˧˥	High rising	“hemp”
3	馬	˨˩˦	Low falling	“horse”
4	罵	˥˩	High falling	“scold”

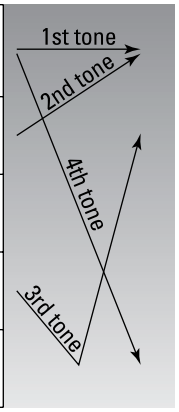


Figure 15-5: The Mandarin tonal system.

Illustration by Wiley, Composition Services Graphics

Check this native production of Mandarin Chinese at `www.utdallas.edu/~wkatz/PFD/Mandarin_tongue_twister.html`.

The tone systems can get even more elaborate. Cantonese has seven tones in Guangzhou and six in Hong Kong. Figure 15-6 shows the six-tone system. When poetry is considered (with entering and departing tones factored in), the tally can reach up to nine tones! You can imagine the fun one can have with Cantonese tongue twisters.

Chinese Character	IPA Tone Symbol	Tone Description	English Translation
詩	˥	High level	“poem”
試	˦	Mid level	“to try”
事	˩	Low level	“matter”
時	˨˩	Low falling	“time”
史	˧˥	High rising	“history”
市	˧˦	Mid rising	“city”

Figure 15-6: The Cantonese tones.

Illustration by Wiley, Composition Services Graphics

Tone sandhi (Sanskrit for *joining*) is a change of tones in tonal languages when some tones are chained next to each other. Not all tone languages have tone sandhi, but many do. Mandarin has a relatively simple sandhi system, yet it's important to know them if you want to sound like a fluent speaker.

Some other facts you should know about tone languages include the following:

- *Tone* (phonemic tone) is when pitch changes meaning in language.
- Many Asian languages (like Chinese, Thai, and Vietnamese) are tonal.
- Approximately 80 percent of African languages are tonal. Hausa, Igbo, Yoruba, and Maasai are common examples.
- In South America, many pre-Columbian languages such as Mayan are tonal.
- Many Amerindian languages are tonal, including over half of the Athabaskan family (including Navajo).
- It's not clear why some regions have tone languages and others don't. Ancient Greek was tonal, and these sounds contributed to the early Greek writing system. However, Modern Greek has lost its tonal quality.
- Linguists have recently discovered an African-style register tone language in Southeast Asia, making the picture even more complex.

Tracking Voice Onset Time

Voice onset time (VOT) refers to the amount of time (measured in milliseconds, or ms) between the release of a stop consonant and the onset of voicing. If you say "pa" and exaggerate the time frame between blowing the lips apart (a gesture that creates an acoustic event known as the *burst*) and the moment that the vocal folds begin to buzz for the /ɑ/, you make a really long VOT.

This time gap is an important cue telling listeners that the initial syllable is voiceless, rather than voiced. That is, people's ears can pick up on that 30 to 80 ms chunk of time and determine that you intend /pɑ/, /tɑ/, or /kɑ/, instead of /bɑ/, /dɑ/, or /gɑ/. If you start voicing at almost the same time as the burst (a short-lag VOT), listeners will hear this as voiced (/bɑ/, /dɑ/, or /gɑ/).

Your VOT values are precisely timed. They vary by place of articulation (for example, bilabial, alveolar, velar), and also by factors particular to each language. These sections cover important differences you can expect as you explore some of the languages of the world.

Long lag: /p/, /t/, and /k/

English voiceless stop consonants are typically about 30 to 50 msec in length. They differ in length based on how much aspiration there is. In many contexts, stop consonants are produced with a burst, but little or no aspiration, such as the [p] in the word "rapid" [ˈɹæpɪd]. In such cases, VOT is typically shorter than when aspiration is present. Figure 15-7 shows waveforms of different voiceless stop consonants to give an idea of how different languages separate voiced from voiceless.

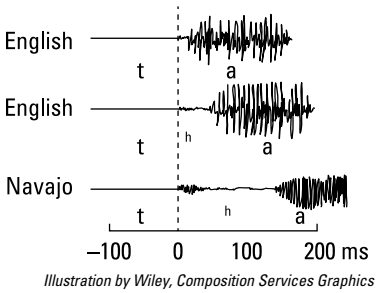


Illustration by Wiley, Composition Services Graphics

Figure 15-7: Waveforms showing VOTs of two voiceless stops from English and one from Navajo.

Figure 15-7 shows two "t"s for English: an unaspirated "t" (as in "stop"), and an aspirated "t" (as in "top"), shown in the middle of the figure. Compare these with Navajo, a language known for its high amount of aspiration. Navajo has VOT values of about 150 ms for its /k/ voiceless stops, which is a really long lag. Listen to this link, where a speaker of Najavo is saying Ke'shmish (Christmas) at `www.utdallas.edu/~wkatz/PFD/Navajo_Keshmish.wav`. For more information about learning Navajo (Dene), including many sound examples, see `http://navajopeople.org/navajo-language.htm`.

Short lag: /b/, /d/, and /g/

How about the voiced side of the spectrum? Most English speakers fall into one of two camps:

- Stop consonants are produced with short VOT values (zero to 20 ms).
- In some cases, stop consonants are produced with negative values (known as *prevoicing,* described in the next section).

Figure 15-8 shows VOT values for English voiced stops compared to different languages. Notice that the English values hover between zero to slightly negative. Spanish and Thai, however, can be much more negative.

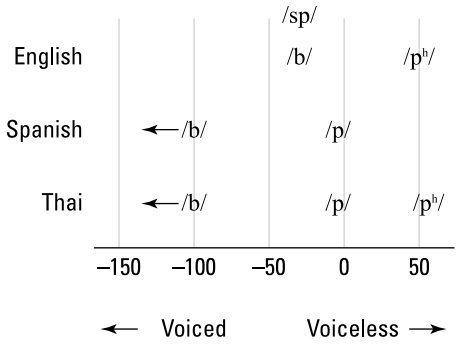


Figure 15-8: VOTs of voiced and voiceless stops in three languages.

Illustration by Wiley, Composition Services Graphics

In the case of Thai, there is a three-way split between the "b" and "p" continuum (whereas English has only a two-way distinction). Take a look at Table 15-1.

Table 15-1 Thai Three-Way Stop Consonant Split

IPA	*English Translation*	*VOT Split*
/pʰɑ/	"cloth"	Voiceless — aspirated
/pɑː/	"aunt"	Voiceless — unaspirated
/bɑː/	"crazy"	Voiced

Look at Figure 15-8 and compare Spanish with English, two languages familiar to many speakers in North America and Europe. Like English, Spanish has voiced and voiceless stop consonants. However, English is often aspirated, while Spanish isn't. Specifically, English voiceless phonemes are aspirated at the beginning of a syllable (such as in the word "peak") and unaspirated elsewhere (such as in "speak" or "hip"). In contrast, Spanish voiceless phonemes are produced without much of a VOT, similar to the case in the English word "speak."

As in English, the Spanish /p/ (as in the word "peso," pronounced /ˈpeso/), is distinct from its voiced counterpart, /b/, as in the word "beso" (kiss), pronounced /ˈbeso/. That is, Spanish and English both make a two-way distinction in voicing. The example given here is for /p/ versus /b/, but this also holds for /t/ versus /d/ and /k/ versus /g/.

Pre-voicing: Russian, anyone?

Pre-voicing is when voicing begins before the stop consonant is released. It's a negative VOT. Some English speakers pre-voice more than others, but overall English voiced stops generally range from slightly negative (–20 ms) to short lag (20 ms) VOTs.

Many linguists consider English to be rather, well, wimpy in the voicing department. According to these folks, in a true voicing language the contrast in word-initial position is between voiceless unaspirated stops and prevoicing. This is shown in the case of Spanish (and also found in Dutch, French, Hungarian, and Russian). In such true voicing languages, voiced stops have strongly negative VOTs. A recent study has shown VOT values of approximately –100 ms for the /d/ in these utterances as in the Russian word "da" (*yes*).

If a language sets a voiced sound to be so negative in VOT, then the voiceless counterpart doesn't have to be strongly voiceless (as in Navajo). For instance, French has a voiced/voiceless, two-way opposition, like English. Similar to Spanish and Russian, French uses very pronounced, pre-voiced VOTs for its voiced sounds. On the other hand, its voiceless utterances are actually produced with short-lag VOTs. Recall that short-lag VOTs for English speakers indicate voiced stops.

This means if a French voiceless phoneme (for example, [t]) was cut out and stuck in English speech, it would likely sound like the voiced phoneme [d]. However, compared to the far negative prevoiced sounds produced for the voiced sounds in French, such short-lag voiceless segments sound just fine. These facts illustrate how different languages use different points along the VOT continuum to form boundaries among stop consonants.

Chapter 16

Visiting Other Places, Other Manners

In This Chapter

- Tuning in to phoneme timing
- Checking out different manners of articulation for familiar places
- Voyaging to new places of articulation you've probably been *too scared to visit!*

Languages can vary from English in more ways than having alternative breath streams or phonemic tone (check out Chapter 15 for more info). Languages can have differences in the length of speech sounds and in the place and manner by which the sounds are produced. Nothing is more fun than exploring the sounds of the world's languages with your very own mouth in the comfort of your living room. So sit back, relax, and get ready for a world tour of language place and articulation, starting now.

Twinning Your Phonemes

Ready for double trouble? In English spelling, doubling a letter usually has no effect on sound. If you listen to the middle consonant in "petting," "running," or "tagging," there's nothing especially long about the /t/, /n/, or /g/ middle sounds. The doubling is usually only for spelling, and these words would be written in the International Phonetic Alphabet (IPA) with a single medial phoneme, such as /t/ in /ˈpɛtɪŋ/ (for "petting"). In other English words called *compounds* (made by combining two stand-alone words), *geminates* (doubled consonant sounds) can be found, such as in "boo*kk*eeper" and "ca*tt*ail." In these compound words, doubling letters isn't only a case of spelling but also results in longer consonant sounds.

To produce a *geminate* (meaning twin), make a consonant articulation and hold it for approximately twice the length as normal. Languages with geminate consonants include Arabic, Finnish, Hungarian, Italian, Japanese, Russian, and Slovak. Check out Table 16-1 for examples in Italian.

Table 16-1 Italian Consonant-Length Contrast Pairs

Manner	*Geminate*	*English Translation*	*Nongeminate*	*English Translation*
Nasal	/kanna/	"rod"	/sana/	"health"
Lateral approximant	/balla/	"she dances"	/kala/	"it subsides"
Fricative	/piɔve/	"it rained"	/piɔvve/	"it rains"

If you're a teacher of English as a second language, geminates are important because this timing difference can sometimes cause interference in English pronunciation.

The actual amount of time a talker spends making a geminate longer varies from language to language. Overall, geminates are usually about 1.5 times as long as regular consonants. What's really important is that geminates sound longer to listeners. Seeming double long is really more in the ear of the listener.

Visualizing vowel length

Consonants aren't the only sounds that can be doubled: *Vowel length* can also play an important role in languages. English is again a linguistic odd man out because vowel length distinctions are fairly common among the world's languages. They can be found in Finnish, Fijian, Japanese, and Vietnamese. Vowel length doesn't work phonemically (at the meaning level) for English speakers. For example, "today" and "tooodaayy" mean the same thing. This isn't true in languages that have vowels that are extra long or extra short.

The IPA method for marking an extra-long vowel is to place a colon-like mark after it [ː]. For extra-short vowels, a *breve* mark (meaning "brief") is placed above the vowel [˘]. Tables 16-2 and 16-3 show more examples.

Table 16-2 Japanese Vowel-Length Contrasts

Regular Vowel	*IPA*	*English Translation*	*Long Vowel*	*IPA*	*English Translation*
kiro	/kiro/	"kilogram"	kiiro	/kiːro/	"yellow"
obasan	/obɑsɑn/	"aunt"	obaasan	/obɑːsɑn/	"grandma"
soshiki	/soʃiki/	"system"	sooshiki	/soːʃiki/	"funeral"

For sound files in Japanese, visit `www.utdallas.edu/~wkatz/PFD/Japanese_vowel_length_contrasts.html`.

Table 16-3 Hausa Vowel-Length Contrasts

Long Vowel	*IPA*	*English Translation*	*Short Vowel*	*IPA*	*English Translation*
d`ashē	/dɑˈʃe/	"seedling"	(a) d`ashe	/ɑdɑˈʃĕ/	"transplanted"
dukā	/dukɑ/	"beating"	duka	/dŭkɑ/	"all"
jīma	/ʤimɑ/	"tanning"	(an) jima	/ɑnˈʤĭmɑ/	"(one has) spent time"

For sound files in Hausa (Nigeria), visit `http://aflang.humnet.ucla.edu/Hausa/Pronunciation/vowels.html#anchor702260`.

Tracking World Sounds: From the Lips to the Ridge (Alveolar, That Is)

Journeys usually start from home, from the more familiar to the less well known. In this articulatory cruise, you begin with sounds made at the front of the mouth and work toward the back.

Looking at the lips

English has a decent number of consonants produced at the lips. These include oral stop phonemes /p/ and /b/ and the nasal stop /m/. However, some other downright fascinating sounds can be produced at this part of the body.

Fricatives

Starting with *fricatives*, the sound /ɸ/ (*phi,* named after the Greek symbol) is produced by moving the lips together as if making a "p" but instead leaving a very slight opening so a hissing sound is made. Because this sound is relatively quiet, it's *marked* (uncommon) in the world's languages. You can find

this sound in many of the Japanese words that are (wrongly) transcribed into English with an "f," such as "Fuji" or "fugu" (if you happen to have a hankering for poisonous blowfish!).

If a labial fricative is voiced, it's transcribed as /β/, the symbol *beta*. You find this sound in Spanish for many words written with a "b," such as "ha*b*er" (to do, make) or with a "v," such as "*v*erde" (green). Actually, Spanish isn't pronounced with a labiodental "f" or "v" but with approximants instead. This is probably why your fourth grade Spanish teacher kept telling you over and over again to watch her and say it the way she does.

Labiodental

Labiodental sounds result where the lips meet the teeth. English has the fricatives /f/ and /v/, as in "*f*at" and "*v*at." The IPA includes a symbol, /ɱ/, for nasal sounds produced at this place of articulation. You produce this "mf" kind of sound in English by saying words where an /m/ and an /f/ sound come together, such as "e*mph*asis." No languages seem to use this sound as a stand-alone phoneme, but /ɱ/ does occur as an *allophone* (context-sensitive variant of corresponding bilabials).

Labiodental approximant

You produce the voiced *labiodental approximant* /ʋ/ (the IPA symbol *script v*) by putting your lips in the position for a "v," but instead of hissing, you bring the lips together like in a "w" motion. Quite a few languages use this sound as types of /w/ allophones. Some languages, such as Guarani, an indigenous tongue in Paraguay, contrast this approximant phonemically with velar and palatal approximants.

Dusting up on your dentals

English speakers commonly say dental fricatives /θ/ and /ð/ in words such as "*th*ick" and "*th*is." Depending on your accent, you may also produce the stops /t/ and /d/ at the teeth, although most North American speakers produce these sounds at the alveolar ridge. There is quite a bit of individual variation.

Dental stops are also produced when a consonant comes before another dental sound, as in "ni*nth*" and "hea*lth*." The symbol for *dentalization* is a small, staple-like diacritic placed under an IPA character [̪]. In English, this kind of variation is due to anticipatory coarticulation (see Chapter 6). For

example, in "ni*n*th" and "hea*l*th," while the /n/ and /l/ are being produced, the tongue is already getting in position for the upcoming /θ/ and thus moves forward to a dental position (instead of the usual alveolar position).

Malayalam is a Dravidian language spoken in southern India by approximately 36 million people. Malayalam is the official language of the state of Kerala, but it is also famous for its nasals! It contrasts nasal stops at six places of articulation, including dental. Table 16-4 gives you some examples.

Table 16-4 Six Places of Nasal Consonants in Malayalam

	Bilabial	*Dental*	*Alveolar*	*Retroflex*	*Palatal*	*Velar*
IPA	[kummi]	[puṉṉi]	[kunni]	[kuɳɳi]	[kuɲɲi]	[kuŋŋi]
English Translation	"shortage"	"pig"	"virgin"	"link in chain"	"boiled water and rice"	"crushed"

Notice also that Malayalam has geminate nasal consonants. You can access sound files by a native speaker at `www.utdallas.edu/~wkatz/PFD/Malayalam_consonants.html`.

Assaying the alveolars

An *alveolar consonant* is a sound produced by restricting airflow at the alveolar ridge, a raised part of your anatomy just behind your upper teeth. Refer to Chapter 2 for more about the alveolar ridge. Many scientists believe the alveolar ridge resulted from strong evolutionary pressures for speech. No matter where this lovely ridge came from, it's clear that alveolar consonants span all manners of articulation. English has a stunning representation of alveolar consonants, including /t/, /d/, /n/, /ɹ/, /ɾ/ (tap), /s/, /z/, and the lateral approximant /l/.

Other interesting alveolar sounds in the IPA chart not represented in English are the lateral fricatives, /ɬ/ and /ɮ/. These alveolar lateral sounds, like English /l/, are made by directing airflow around the sides of the tongue. However, in the case of these fricatives, you hiss instead of just approximating (as in a /w/ or /j/). The voiceless alveolar lateral /ɬ/ is fairly common and found in Welsh, Navajo, Taiwanese, Icelandic, and Zulu. The voiced phoneme /ɮ/ is rare, although Zulu contrasts voiceless and voiced alveolar laterals. Check out these alveolar lateral examples for your next visit to KwaZulu-Natal in Table 16-5.

Table 16-5 Some Zulu Lateral Alveolar Sounds

Lateral	*IPA Symbol*	*Word*	*Spelling*	*English Translation*	*Pronunciation*	*Sound Files*
Approximant	[l]	[lá:là]	lálà	"sleep"	as in English "leaf"	www.utdallas.edu/~wkatz/PFD/Zulu_Lala1.wav
Voiceless Fricative	[ɬ]	[ɬá:la]	hlálà	"sit" or "live"	as in Welsh "Llanelli"	www.utdallas.edu/~wkatz/PFD/Zulu_Lala2.wav
Voiced Fricative	[ɮ]	[ínàɮàlà]	inldlala	"hunger"	Voiced form of [ɬ]	www.utdallas.edu/~wkatz/PFD/Zulu_Lza.wav

Flexing the Indian Way

From a culture that brought the world yoga, it stands to reason that the fascinating property of retroflex would emerge from the Indian subcontinent. You produce the retroflex sounds with the tongue curled back toward the rear of the mouth such that a slightly post-alveolar region of the palate is the point of articulation. See Figure 16-1 for a diagram showing the tongue in retroflex position.

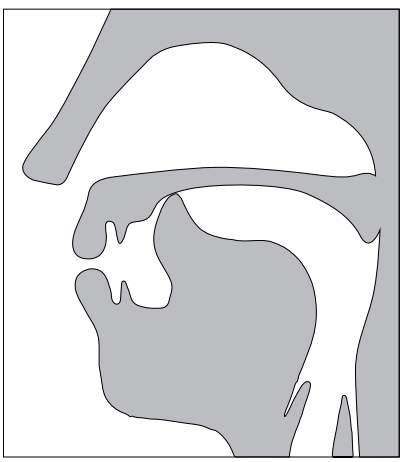


Figure 16-1: Producing retroflex.

Illustration by Wiley, Composition Services Graphics

Retroflex is both a *manner* (shape of tongue) and *place* (region of the palate) feature. You transcribe these sounds in the IPA by placing a hook diacritic on the bottom right of the symbol. The retroflex consonants are common in Indian and Pakistani English and can be easily exemplified in an Indian English pronunciation of a famous food of the region — curry rice /ˈkʊɽi ɽaɪʂ/. Indeed, a range of sounds may be produced as retroflex, as shown in Table 16-6.

Table 16-6 List of Retroflex IPA Symbols

	Nasals	*Plosives*	*Fricatives*	*Approximants*	*Flap*	*Lateral Approximants*
IPA	/ɳ/	/ʈ/ /ɖ/	/ʂ/ /ʐ/	/ɻ/	/ɽ/	/ɭ/
Sample Languages	Malayalam, Kannada, Vietnamese	Bengali, Hindi, Telegu	Mandarin, Russian, Slovak	Malayalam, Pashto, Tamil	Hindi, Hausa, Japanese	Kannada, Malayalam, Swedish

To make a retroflex "r," begin by producing a familiar English alveolar /ɹ/ in a VCV (vowel-consonant-vowel) context, /ɑɹɑ/. Following Figure 16-1, flap your tongue back so you make a kind of hollow sound during the /ɻ/, to produce /ɑɻɑ/. You should be releasing your inner Indian! If you have had luck with this, try it with an "s" /ɑʂɑ/, then an "n" for /ɑɳɑ/. You can also follow these spoken examples by a native speaker of Hindi:

- /ɑɻɑ/ (`www.utdallas.edu/~wkatz/PFD/Hindi_ara.wav`)
- /ɑʂɑ/ (`www.utdallas.edu/~wkatz/PFD/Hindi_asa.wav`)
- /ɑɳɑ/ (`www.utdallas.edu/~wkatz/PFD/Hindi_angra.wav`)

Passing the Ridge and Cruising toward the Velum

In this section, you discover the region in the middle of your mouth. This midmouth region includes the post-alveolar (also called palato-alveolar) and palatal regions. Anatomically, it's the terrain of the hard palate, a relatively solid zone of the roof of your mouth with underlying bone. Here I provide more details about how consonant sounds are made in this region by talkers of the languages of the world.

Studying post — alveolars

English has two post-alveolar fricatives, /ʃ/ and /ʒ/, and two affricates, /tʃ/ and /ʤ/. These are produced at roughly the same part of the mouth as retroflex consonants, although with a very different tongue position. Retroflex consonants (see the earlier section "Flexing the Indian Way") have a hollow tongue shape, whereas post-alveolars have a humped shape. Another way of saying this is that retroflex sounds are *apical* (made with the tongue tip), while post-alveolars are *laminal* (made with the tongue blade). I describe it further in the "Working with Your Tongue" section later in this chapter.

Populating the palatals

Palatal consonants are sounds produced by constricting airflow at the hard palate. English has just one lonely palatal consonant, the approximant /j/, as in the word "*y*ellow." However, other languages have different manners of sounds (including stops, nasals, fricatives, and approximants) produced at the palatal place of articulation. Here is a sampling:

- **Voiceless palatal stops:** The letter "c" stands for a voiceless palatal stop in the IPA. It sounds like a "k" but is produced slightly more forward. To make this sound, try making a familiar English glide /j/, but at the same place of articulation produce a stop. Try /aja/, then /aca/. After you get them down, you'll be able to say *red* in Albanian ([kuc]) and *sack* in Macedonian ([ˈvrɛca]).
- **Voiced palatal stops:** If a palatal stop is voiced, it's written in IPA like an upside-down "f." It sounds like a fronted or partially palatalized /g/, as in the English word "argue." Voiceless and voiced palatal stops are found in Basque, Czech, Dinka, Greek, Irish, Slovak, and Turkish.
- **Palatal nasals:** Written like an "n" without a left hook ([ɲ]), they're found commonly in Spanish, in such words as "peña," "señor," and "año." ***Note:*** Although Spanish writing uses a tilde "˜" character over the "n" for these sounds, this is just for spelling and not for the IPA.
- **Voiceless palatal fricatives:** The sounds /ç/ and /ʝ/ strike the ear much like the English fricatives /ʃ/ and /ʒ/, but they're produced slightly farther back in the vocal tract. The voiceless /ç/ is found in many varieties of German, in words such as "Ich" (*I*) and "nicht" (*not*).
- **Voiced palatal fricatives:** The voiced palatal fricative /ʝ/ is a rare sound, occurring in only 7 of the 317 languages surveyed by the UCLA Phonological Segment Inventory Database (UPSID).
- **Palatal lateral approximants:** These sounds are produced similar to making an English (velar) *dark l*, although they're slightly fronted to the palatal place. Languages that have lateral approximant consonants include Basque, Castillian Spanish, Greek, Hungarian, Norwegian, and Quechua. Italian offers a good example, as seen in Table 16-7.

Table 16-7 Some Italian Palatal Lateral Approximants

Example	*IPA*	*English Translation*	*Sound Files*
figlio	[ˈfiʎʎo]	"son"	www.utdallas.edu/~wkatz/PFD/figlio.wav
foglia	[ˈfoʎʎa]	"leaf"	www.utdallas.edu/~wkatz/PFD/foglia.wav

(Re)Visiting the velars

The *velars* are sounds made by blocking airflow at the soft palate and have several categories:

- **Velar stop consonants:** As an English speaker, you use the oral stops /k/ and /g/ and the nasal stop /ŋ/. With /ŋ/, this nasal sound is only permitted at the end of syllables in English in words such as "sing," "sang," and "sung."
- **Voiceless velar fricatives:** Velar fricatives are common in languages throughout the world. The voiceless velar fricative is written in IPA as /x/, as in Johann Sebastian Bach /bɑx/ or Spanish "hijo" (*son*) /ˈixo/.

  This sound is pretty easy to make for English speakers:

  1. **Produce the regular velar stop in the syllable /**ko**/.**
  2. **Try again with a bit more air pressure and your tongue body lowered a tad.**

     You should feel a throat-tickling sensation back where the /k/ air stoppage usually takes place. You've produced the /x/ of Spanish [ˈixo] (son).
- **Voiced velar fricatives:** You can produce the voiced velar fricative /ɣ/, represented by the Greek letter *gamma.* True forms of this sound are found in a number of world languages, including Arabic, Basque, Greek, Hindi, Navajo, and Swahili.
- **Velar approximants:** A close cousin of the voiced velar fricative /ɣ/ is the velar approximant /ɰ/. This rather odd-looking character indicates a velar articulation that's not quite as closed as a velar fricative. In a way, it's a lowered velar fricative. The phoneme /ɰ/ is found in some Spanish words spelled with "g," such as "diga" /ˈdiɰɑ/, ([you] *speak*) and "pago" /ˈpɑɰo/ ([I] *pay*). ***Note:*** There are some stylistic differences in transcribing spelled "g" in Spanish, with some phoneticians preferring to use /ɣ/ and others noting that /ɰ/ is usually more correct.
- **Velar lateral approximants:** Small capital "L" is reserved in the IPA to represent the relatively marked (unusual) velar lateral approximant.

IPA symbol /ʟ/ represents a voiced sound, although even rarer voiceless varieties have also been reported. Two things can be learned by the beginning phonetician about /ʟ/ at this point:

- You can use IPA /ʟ/ to transcribe Mid-Waghi, a Trans-New Guinean Language of Papua New Guinea with approximately 100,000 speakers.
- You can't use IPA /ʟ/ to transcribe the word "Larry" in English (see Chapter 20). If you do so, your phonetics instructor has permission to extradite you to central New Guinea.

Heading Way Back into the Throat

For some rather understandable reasons, many English speakers don't like to produce speech at the very back of the throat. This probably results from upsetting memories of dental visits or childhood fears of swallowing really hot beverages, but one thing is certain: such bad experiences can prevent you from producing sounds that much of the world enjoys. In this section, I lead you into the dark recesses of your vocal tract to experience bold new vocal horizons.

Uvulars: Up, up, and away

Uvular stops are found commonly in the Semitic languages, including (Sephardic/Mizrahi) Hebrew and Arabic. This is why common Arabic words thought to begin with a "k" sound are often spelled in English with the letter "q" (*Quran*, *Al-Qaeda*). A *uvular stop* is a constriction of airflow involving the uvula, the dangling part of the soft palate in the back of the throat. The truth is, these words aren't produced with a (velar) "k" but with a stop made farther back in the uvular region. This sound is also found in Quechua (South America), Tlingit, and Aleut (Aleutian Islands, Alaskan region). An example from Aleut is "gaadan" (/ˈqaːðn/), which mean *dolly varden,* a type of fish.

If uvular stops are voiced, they're represented in the IPA as /ɢ/, but you don't use this symbol to transcribe an English word, such as "Greg." You might use it for Yemeni Arabic or Tlingit, such as [ɢuːtʃ], which means *wolf.* To practice other common Tlingit words, check out this instructional site from the University of Alaska Southeast at `www.youtube.com/watch?v=gr-x6EL39PY`.

The next sounds to enjoy are the uvular fricatives: /χ/ and /ʁ/. The voiceless fricative sounds (/χ/) aren't found in English, although they're found in French and German as well as many dialects of Dutch, Swiss German, and Scots. Scots (also known as in Lowland Scots) is a Germanic language spoken in Lowland Scotland and parts of Ulster, Northern Ireland. Here are some examples:

Language	*Word*	*IPA*	*English Translation*
French	proche	[pχɔʃ]	"nearby"
German	dach	[daχ]	"roof"
Scots	nicht	[nɪχt]	"night"

In addition, you can find voiceless uvular fricatives in languages from other families, including Arabic, Haida, Hebrew, and Welsh.

The voiced fricative /ʁ/ is found in French "rouge" /ʁuz/ (*red*) and "rose" /ʁoz / (*rose*) . Many languages have this sound, including German, Hebrew, Kazan, Malay, Tatar, Uzbek, Yiddish, and Zhang.

A good way to make a uvular fricative is to begin with a voiceless velar fricative (see the earlier section in the chapter "(Re)Visiting the velars") and then move backward to your uvular area. You can already produce a /x/, as in Johann Sebastian Bach, right? Just make a hissy sound in the throat, but farther back like this:

1. **Begin with /**bax**/ ("Bach").**
2. **Try to produce /**baχ**/.**

   The hissing should be back at your uvula, the very top posterior of your throat.
3. **Try some Scots, /**nɪχt**/ ("night").**

   Congratulations! You have made a voiceless, uvular fricative!

The IPA also lists /ɴ/ and /ʀ/ in the uvular place of articulation. These symbols represent a uvular nasal and trill, respectively. (See "Going for Trills and Thrills" for more info on trills.) A velar nasal is found in Inuit and Japanese. For example, the Japanese word for "Japan" (Nihon) [nʲihoɴ]. Listen to it at `www.utdallas.edu/~wkatz/PFD/Nihon.wav`.

A uvular trill, /ʀ/, is made in place of voiced uvular fricatives in many languages. You can find more information on uvular trills in the section on manner ("Going for Trills and Thrills").

Finding your uvula

Time to find out where your uvula is. Begin by making a "k" flanked by the back vowel /u/:

1. **Say /**uku**/.**
2. **Drop your chin way down and bring your tongue back.**
3. **Make the "k" farther back in the mouth.**

   It will have a more "throaty" quality. Congratulations, you have said /uqu/.

Try some different vowel contexts: /iqi/ and /aqa/. See if you can pronounce /q/ in syllable-initial position /qaf/, which is the 21st letter of the Arabic alphabet (`www.utdallas.edu/~wkatz/PFD/qaf.wav`).

Pharyngeals: Sound from the back of the throat

The *pharynx* is the back of the throat, commonly known as the throat wall — that's the area that the doctor swabs when you're being checked for strep throat. This part of the vocal tract is constricted for the production of fricatives and achieved by pulling the tongue body up toward the pharyngeal wall. Pharygneal fricatives can be voiceless /ħ/ or voiced /ʕ/. They're considered perfectly nice sounds in languages that have them in their inventory. Table 16-8 shows you some examples of the voiceless /ħ/.

Table 16-8 Examples of Voiceless Pharyngeal Fricatives

Language	*IPA*	*English Translation*	*Sound File*
Hebrew (Sephardic)	[ħaʃˈmal]	"electricity"	`www.utdallas.edu/~wkatz/PFD/Hashmal.wav`
Somali	[ħoːd]	"cane"	`www.utdallas.edu/~wkatz/PFD/Hood.wav`

You articulate the pharyngeal fricative /ʕ/ with the root of the tongue up against the pharynx, but it's voiced. Although called a fricative, this sound is often made with an approximant manner, and no language makes a phonemic distinction between pharyngeal fricatives and approximants. Table 16-9 gives you some examples, including Chechen, a Caucasian language spoken by more than 1.5 million people.

Table 16-9 Examples of Voiced Pharyngeal Fricatives

Language	*IPA*	*English Translation*	*Sound File*
Chechen	[ʕan]	"winter"	www.utdallas.edu/~wkatz/PFD/winter_chechen.wav
Somali	[ʕaːdi]	"normal"	www.utdallas.edu/~wkatz/PFD/somali_caadi.wav

Going toward the epiglottals

Until fairly recently, pharyngeals were thought to be the extreme. Researchers have since realized that in certain dialects of Arabic and Hebrew people produce fricatives at the epiglottis, which is quite a phonetic feat because the epiglottis is the flap located just above the larynx. The chief purpose of the epiglottis is to assist in swallowing and to prevent *aspiration*, which is foreign bodies entering the vocal folds, trachea, or lungs. To produce speech sounds there is, well, impressive.

Semitic languages (such as Arabic, Hebrew, and Aramaic) can have quite a bit of variation between pharyngeal and epiglottal articulation, depending on dialect and individual-talker variability.

The IPA character for the voiced epiglottal fricative is written like a pharyngeal fricative but with a bar through it (/ʢ/). The voiceless epiglottal fricative is denoted with a character like a small capital H (/ʜ/). You can imitate these if you wish.

Table 16-10 provides some examples (with links to sound files) that demonstrate the voiced epiglottal sound (/ʢ/):

Table 16-10 Examples of Voiced Epiglottals

Language	*IPA*	*English Translation*	*Sound File*
Arabic (certain dialects)	[tɑʢɑʃʃæː]	"to have supper"	www.utdallas.edu/~wkatz/PFD/Arabic_tachasshe.html
Hebrew (Sephardic)	[naʢor]	"make a donkey sound"	www.utdallas.edu/~wkatz/PFD/Hebrew_make_a_donkey_sound_sephardic.html

Although the Semitic languages don't have meaningful contrasts between words containing pharyngeal and epiglottal sounds, other languages do. Table 16-11 lists some examples from Aghul, an endangered language in Dagestan (Russia and Azerbaizhan):

Table 16-11 Examples of Pharyngeal and Epiglottal Contrasts

IPA	*English Translation*	*Sound File*
/mɛʜɛr/	"wheys"	`www.utdallas.edu/~wkatz/PFD/Aghul_mehar.aiff`
/muħar/	"barns"	`www.utdallas.edu/~wkatz/PFD/Aghul_muhar.aiff`

It can be difficult for native English speakers to constrict the pharynx for Arabic and other Semitic language sounds. However, you can master it after a lot of practice. One way is to just try and imitate native speakers.

Here are some tips from Europeans trying to learn Arabic:

- **Gag.** You'll feel the muscles of your throat constrict the passage of air in basically the right way.
- **Voice the sound.** This means that your vocal cords vibrate when making it. It sounds like the bleating of a lamb, but smoother.
- **Act as if you're being strangled while you're swallowing the "ah" sound.** This tip comes from a world expert in colloquial Egyptian Arabic.

Please note these scary-sounding tips are just for the beginning. After these sounds are realized, they can be produced easily and there's nothing scary about them.

Working with Your Tongue

The tongue has different functional regions, including the tip (apex), blade, middle, and back. Most of the action of the tongue is in a front–back direction, although shaping the tongue's sides is also important to distinguish liquid ("r" and "l") sounds and fricatives, including /s/ and /ʃ/.

Sounds made with the tongue tip or blade are called *coronal* (meaning the crownlike upper portion of a body part) sounds. Coronal is an important natural class in phonetics and a functional grouping that distinguishes sounds found throughout the languages of the world. Coronal sounds are made with

the tongue tip or blade raised toward the teeth, alveolar ridge, or hard palate, such as /s/, /t/, /n/, /θ/, and /ð/.

The world of coronal sounds can be further divided into the tongue tip and the tongue blade. Although it may seem confusing, the tip and the blade provide a good opportunity to see how different types of phonetic concepts can be applied to language sounds. Because retroflex consonants are produced with the tongue tip raised (such as Indian English /ɳ/, /ʂ/, or /ɻ/, among others), they're apical. In contrast, post-alveolar consonants such as /ʃ/ and /ʒ/, as in "ship" and "leisure," are produced with a humped tongue blade and are laminal. Although some phoneticians stress the place of articulation differently (retroflex versus post-alveolar), other phoneticians consider them all post-alveolar and specify only the parts of the tongue involved.

Table 16-12 may help you with understanding this concept.

Table 16-12 Classifying Coronal Sounds in the IPA

Example	*IPA*	*Classification System 1 (by place of articulation)*	*Classification System 2 (by tongue region)*
(Malayalam) bhaasa "language"	/ˈbʰaːʂa/	Retroflex	apical
*"sh*e, lei*sur*e"	/ʃ/, /ʒ/	Post-alveolar	laminal

Going for Trills and Thrills

A *trill* is a consonant made by allowing an articulator to be repeatedly moved under air pressure. Whereas a tap strikes the articulatory region only once, a trill usually vibrates for two to three periods and sometimes up to five. A good example to keep in mind is what people commonly call the *rolled r* of Spanish, in a word like "burro" (*donkey*) or "perro" (*dog*).

Most speakers of English don't produce trills, although they're found in many other common Indo-European languages, including Spanish, Czech, French, Polish, Russian, and Swedish. Trills are found in some varieties of English, including Scottish English.

Table 16-13 shows the trills listed in the IPA, along with some languages that have them. Notice that trills can occur at different places of articulation. Bilabial trills (denoted with the IPA symbol /ʙ/) are relatively rare, reported chiefly in some Austronesian languages, like Kele. Coronal and uvular trills are more common.

Table 16-13 Various Kinds of Trills

Trill	*Symbol*	*Language*	*Spelling*	*IPA*	*English Translation*	*Sound File*
Bilabial	/ʙ/	Kele	N/A	/mʙuɛŋkeiʔ/	fruit	www.utdallas.edu/~wkatz/PFD/Kele-fruit.aiff
Coronal	/r/	Polish Spanish	krok, oro	[ˈkrɔk], [ˈoro]	step, gold	www.utdallas.edu/~wkatz/PFD/Polish_krok.wav
Uvular	/ʀ/	French (some dialects), German	rendez-vous, rübe	[ʀɑ̃devu], [ˈʀyːbə]	appoint-ment, turnip	www.utdallas.edu/~wkatz/PFD/Fr-Rendez-vous.wav

To make a trill, you set an articulator in motion by having it move under air pressure. The moving articulator can be the lips (for a kind of raspberry-like effect), the tongue blade (for the trilled /r/ as in Spanish "bu*rr*o"), or the uvula (the hanging part of the roof of your mouth, way in back).

Ready to try some trills? The alveolar trilled /r/ isn't too difficult. Follow these steps:

1. **Make a conventional English /ɹ/ in the VCV context, /ɑɹɑ/.**
2. **Allow your tongue to roll as the "r" is produced.**

   If the trilling isn't happening, keep your mouth more open.
3. **Relax and have your mouth open by imagining you have a pencil held between your teeth.**
4. **If this doesn't work, try placing a real pencil (eraser side in!) between your teeth for spacing, then try again.**

   You sound Spanish, no?
5. **Make a trill way back there.**

   To make the uvular trill, /ʀ/, you'll be making your uvula jiggle a few times. This might sound a bit extreme, and if you aren't used to these sounds, you may actually think about clearing your throat. That will at least get you to the right neighborhood.

Prenasalizing your stops or prestopping your nasals

Some African languages spell words with an "m" before "b," as in "Mbeke," or an "n" before "d" as in "Ndele," because the sound systems in these languages have *prenasalized consonants* — a nasal and a consonant produced together as one phonetic unit.

To produce a prenasalized stop, you make an oral closure while lowering the velum (that is, opening the nasal passageway). Then you produce a short nasal consonant, followed by a velar raising and an oral release, resulting in an oral stop. This results in a sound that has both nasal and oral qualities, starting with the nasal (slightly) first. These sounds are found among Bantu languages of Africa (Swahili), in Papua New Guinea (Tok Pisin), as well as in Melanesia (Fijian). The easiest way to make these sounds is to follow examples, such as these words in Swahili, in Table 16-14.

Table 16-14 Some Prenasalized Stops in Swahilli

Word	*IPA*	*English Translation*	*Sound Files*
Ndio!	/ndio/ . . .	Yes! (I speak Swahili.)	www.utdallas.edu/~wkatz/PFD/ndio_ninazumgumza_kiswahili.wav
Ndimu	/ndimu/	lemon	www.utdallas.edu/~wkatz/PFD/Swahili_lemon.wav
Mbali	/mbɑli/	far	www.utdallas.edu/~wkatz/PFD/mbali.wav

Talkers can engage in the oral stopping and nasalization processes in the opposite order and produce stops with a nasal release. In these gestures, an oral stop is made just slightly before a nasal. You can find these sounds, just like the prenasalized stops, for *homorganic consonants* (same place of articulation). The combinations /bm/ and /dn/ are examples in English. They occur in English sound combinations like "clu*bm*an" and "gla*dn*ess."

English also has a phonological rule that permits homorganic stop/nasal consonants to be released into the nasal cavity instead of the usual oral release. For example, the word "ridden" is usually pronounced [ˈɹɪdⁿn̩]. The diacritics (the little, fine symbols) used here indicate that the /d/ isn't released orally and the /n/ is syllabic (see Chapter 6 for more details).

Unlike English, many Slavic languages can have nasal consonants that are produced with an audible release even when they begin a word, such as in the name of the *Dn*iester River. These sounds are called *prestopped nasals* because phoneticians think that through historical processes these special sounds resulted from a very short stop consonant (for example, /b/ or /d/) being inserted before a nasal or lateral (such as /m/, /n/, or /l/). For this reason, some phoneticians transcribe them as /ᵈn/ and /ᵇm/, showing the (oral) stopping with a small diacritic on the left. Phonetically, these prestopped nasals are similar or equivalent to stops with a nasal release (as found in English such as the word "hidden"). However, phonologically (in terms of the rule systems of language) prestopped nasals stand on their own as a single, independent phoneme. Chapter 5 explains allophones and phonemes in more detail.

Table 16-15 shows some examples of Russian prestopped nasal consonants.

Table 16-15 Russian Examples of Prestopped Nasal Words

Word	*IPA*	*Sound Files*
Dniester (River)	[ᵈnistər]	www.utdallas.edu/~wkatz/PFD/Russian_Dniester.wav
day	[ᵈnjom]	www.utdallas.edu/~wkatz/PFD/Russian_day.wav

Rapping, tapping, and flapping

A *tap* is a rapid, single stroke of an articulator. It is a very quick stop, made without time for a release burst to take place. English has the well-known alveolar tap (/ɾ/). This sound is quite common as an allophone of /t/ and /d/ in North American English (see Chapters 8 and 9) and also occurs as an allophone of /ɹ/ in some dialects such as Scottish ("pearl" pronounced as [ˈpɛɹl̩]).

Advancing your tongue root

Phoneticians are ever on the prowl for new sound distinctions in language. As information comes in on newly discovered sound systems, it sometimes becomes necessary to resort to a new feature. One such case is *advanced and retracted tongue root (ATR/RTR),* which are languages with vowel systems that differ based on whether the pharyngeal cavity is expanded or not. The languages that led to this distinction are mainly in West Africa (for example, the Akan language of Ghana), but they're also found in Kazakhstan and Mongolia.

Flap or tap? Tomato or tomahto

If you're a really hardcore phonetics junkie, you probably want to know the difference between a *tap* and a *flap*. Although the average person probably doesn't lose sleep over this question, it has bugged some phoneticians for years. Technically, a tap is when the tongue goes up to an articulation and comes back down again in the same direction.

In English, the alveolars seem to most fit this bill. By this same logic, a flap is defined as when the tongue goes up, hits (tangentially), and then follows through in a, well, flapping motion. From back to front and carrying on. This seemed to describe everything else. However, because the difference between a flap and a tap does not change meaning in English, the terms flap and tap are often used interchangeably. Also, no language contrasts a flap and a tap at the same place of articulation. Therefore, to keep things simple, I refer to any rapid stop-like articulation in this general manner class as a tap (following most conventions in phonetics instruction).

People who make vowels with *Advanced Tongue Root (+ATR)* move the tongue root forward and expand the pharynx (and often lower the larynx), causing a differing vowel quality, including added breathiness. To indicate such a vowel, the IPA uses a *small pointer diacritic* (called *left tack*), which looks like a pointer arrow on a keyboard. This diacritic is placed under the vowel symbol.

In vowels that are *Retracted Tongue Root* (*RTR,* also known as *–ATR*), the tongue root either stays in a neutral position or is slightly retracted. A retracted tongue root is indicated in IPA with a small right tack diacritic placed beneath the vowel symbol. Figure 16-2 shows this distinction from studies of Igbo, a West African language. This figure shows the vocal tract of a talker whose tongue is in the Advanced Tongue Root (solid line) and Retracted Tongue Root (dotted line) positions.

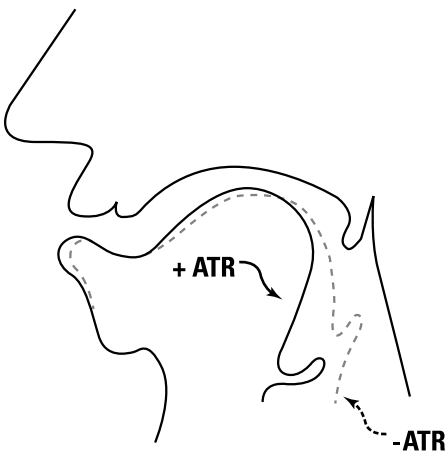


Illustration by Wiley, Composition Services Graphics

Figure 16-2: Comparing +ATR and -ATR in Igbo.

Phonetician Peter Ladefoged and colleagues have done pioneering work with X-ray cineradiography of speakers producing vowels with +ATR/-ATR contrasts. In Table 16-16, you can see a *minimal pair* from Akan from the UCLA phonetics lab website. I provide URLs to sound files in the third column so that you can hear the differences between +ATR and -ATR vowels.

Table 16-16 Akan Vowels That Differ in ATR/RTR

Example	*IPA*	*Sound Files*
"break"	/bu̟/	`www.phonetics.ucla.edu/appendix/languages/akan/a3.aiff`
"get drunk"	/bu̠/	`www.phonetics.ucla.edu/appendix/languages/akan/a4.aiff`

If you wish to speak Igbo or Maa (Maasai), you need to start working on your ATR +/- vowel contrasts. Maasai teachers call the +ATR vowels "close" and the -ATR vowels "open". You can find a nice listing of the Maasai contrasting tongue root vowel sets, with practice words and audio files, at `http://darkwing.uoregon.edu/~maasai/Maa%20Language/maling.htm`.

Phonemic nasalization: Making your vowels nasal for a reason

An English vowel becomes nasalized when it precedes a nasal consonant. An example is "fate" [fet] versus "faint" [fẽnt]. This effect is contextual and goes by various names. Phonologically, it is called *assimilation*, one sound becoming more like another. It is also a kind of *coarticulation,* where one sound is produced at the same time as another. Here is how you do it: At the same time as (or before) the vowel is being produced, the nasal port is free to open, resulting in a nasalized vowel. See Chapter 8 for more information on assimilation and coarticulation processes.

In English, talkers don't freely produce nasalized vowels without a nasal consonant following. That is, one doesn't find just /fẽ/or /sã/. However, in many languages nasalized vowels *can* stand alone and have phonemic meaning. Examples include Cherokee, French, Gujurati, Hindi, Irish, Mandarin, Polish, Portuguese, Vietnamese, and Yoruba.

Portuguese has a well-known series of nasalized vowels. Because Portuguese has a rich vowel system (including diphthongs, triphthongs, and vowels that alternate pronunciations whether stressed or unstressed), the total number of vowels and diphthongs that are nasalized remains debatable among linguists. According to one system, the nasalized monophthongs can be grouped in this list of five words ("cinto," "cento," "santo," "sondo," and "sunto"). In Table 16-17, I include sound files from a native speaker from São Paolo, Brazil.

Table 16-17 Nasalized Vowels in Brazilian Portuguese

Word	*IPA*	*English Translation*	*Sound Files*
cinto	[sĩⁿtʊ]	"belt"	www.utdallas.edu/~wkatz/PFD/Cinto.wav
sento	[sẽⁿtʊ]	"I sit"	www.utdallas.edu/~wkatz/PFD/Sento.wav
santo	[sɐ̃ⁿtʊ]	"saint"	www.utdallas.edu/~wkatz/PFD/Santo.wav
sondo	[sõⁿdʊ]	"I probe"	www.utdallas.edu/~wkatz/PFD/Sondo.wav
sunto	[sũⁿtʊ]	"summed up"	www.utdallas.edu/~wkatz/PFD/Sunto.wav

Speaking of São Paulo, this is a great opportunity to try a nasalized diphthong without a following nasal consonant. Here you can work on making the nasalized diphthong for "São" in "São Paulo."

1. **Say /saʊ/ (without nasalization).**
2. **Raise the diphthong a bit to get /sɐʊ/.**
3. **Make a nasalized /ɐ̃ʊ̃/ by saying "sound,", and feel it in your nose.**
4. **Try just the /ɐ̃ʊ̃/ by itself.**
5. **Put it together, to make /sɐ̃ʊ̃/.**

   If you need help, listen to this sound clip by a native speaker at www.utdallas.edu/~wkatz/PFD/sao_paolo.wav.

Classifying syllable-versus stress-timed languages

Every language seems to have its own rhythm. This has provided comedians with many opportunities, such as Sid Caeser's rhythm-based spoof of French, German, Japanese, and Russian. You can see this spoof at www.utdallas.edu/~wkatz/PFD/caeser_faux_language_montage.wmv.

Knowing about the rhythmic structure of languages is important in language instruction because these patterns can greatly affect a learner's accent. Phoneticians have described timing commonalities between languages, such

as the stress-timed and syllable-timed language distinction. In stress-timed languages, stress is assigned based on syllable structure. A heavy syllable attracts stress. Heavy syllables are syllables that are loaded up with consonants, such as CVC, CCVC, CCVCC, and so forth. Here, "C" means consonant and "V" means vowel. Therefore, a CVC syllable would be a word like "bit" (consonant-vowel-consonant). A light syllable would be V or VC. Take a look at the English words, noting where the heavy syllable is located.

Example	*Syllable Structure*	*IPA*
*fris*king	CCVCC.VC	/ˈfɹɪskŋ/
un*placed*	VC.CCVCC	/ənˈplest/

If you had to imitate the sounds of these words in nonsense syllables, they would sound like "*dah* da" (for "frisking") and "da *dah*" (for "unplaced"). Alternating loud and soft syllables correspond with other timing units known as metrical feet. You can find a good discussion of metrical feet in *Linguistics For Dummies* by Rose-Marie Dechaine, Strang Burton, and Eric Vatikiotis-Bateson (John Wiley & Sons, Inc.).

In contrast to English, languages such as Spanish have relatively simple syllable structures (mostly CV) and don't base their word stress on the presence or absence of a heavy syllable. These languages have a much more regular (rat-a-tat-tat) timing. Expressed in nonsense syllables, phrases would sound much more like "da da da da" than "da *dah* da *dah*." This is called a *syllable-timed* pattern.

Making pairs (the PVI)

Although the stress-timed and syllable-timed labels have intuitive appeal, phoneticians need a way to put a more precise number on this distinction. One way to judge how stress-timed or syllable-timed a given language is, is to measure how much timing varies systematically in that language. Researchers determine a unit of set durational length in a language (say, vowel length) and then measure how much this durational chunk varies as you move from one syllable to the next. The result is a *pairwise variability index* (PVI), a measure of language timing.

Table 16-18 shows some PVI data for some common world languages.

Table 16-18 PVI Values

Language	*Normalized PVI*
Thai	65.8
Dutch	65.5
German	59.7

Language	*Normalized PVI*
British English	57.2
Tamil	55.8
Malay	53.6
Singapore English	52.3
Greek	48.7
Welsh	48.2
Rumanian	46.9
Polish	46.6
Estonian	45.4
Catalan	44.6
French	43.5
Japanese	40.9
Luxembourg	37.7
Spanish	29.7
Mandarin	27.0

Notice the result is quite a mix in terms of the geography and ethnicities. The languages at the top of the list (Thai, Dutch, German, Tamil, and British English) are languages in which vowel variability is relatively large. These are languages typically called *stress-timed.* In contrast, the languages at the bottom of the list (Luxembourg, Spanish, and Mandarin) have small PVI values and tend to have no stress on any particular words and are called syllable-timed.

If you're a formula person and enjoy measuring language details on your own, the formula for calculating the PVI of any given language is

$$rPVI = \left[\sum_{k-1}^{m-1} \left| d_k - d_{k+1} \right| / (m-1) \right]$$

where m is the number of items in an utterance and d_k is the duration of the k^{th} item. This formula has also been modified for vowels with different durations, called the normalized PVI:

$$nPVI = 100 \times \left[\sum_{k-1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

You can also go to `www.nsi.edu/~ani/npvi_calculator.html` for an online PVI calculator to help you with the computations.

Kenneth L. Pike: Portrait of the field linguist as a young man

In this day and age of everything being *neuro*-this and *cognitive*-that, the achievements of the great linguists and anthropologists tend to be somewhat left in the dust. However, phonetics owes a great debt to Kenneth Lee Pike, a remarkable American teacher and researcher who has given phonetics some important definitions and practical field techniques.

Pike was born in Connecticut and studied theology in Massachusetts with the intention of being a missionary. He then studied linguistics at the *Summer Institute of Linguistics,* at which time he also traveled to Mexico to learn Mixtec. Pike next attended the University of Michigan to earn his Ph.D. under Edward Sapir (best known for the Sapir-Whorf hypothesis, which is that the structure of language affects the perception of reality of its speakers).

Pike swiftly rose in academia, becoming the president of the Summer Institute of Linguistics (now SIL International) from 1942 to 1979. He was also chair of the University of Michigan Linguistics Department from 1975 to 1977 and director of the English Language Institute at the University of Michigan at the same time.

Taking a somewhat different path than his mentor, Pike specialized in phonetics and phonology, general linguistics, and foreign language teaching. Pike wasn't just an armchair linguist. He personally carried out studies of over 100 indigenous languages in the field, including those in Australia, Bolivia, Ecuador, Ghana, Java, Mexico, Nepal, New Guinea, Nigeria, the Philippines, and Peru.

Based on his findings, Pike developed a "unified theory of the structure of human behavior" along largely behaviorist grounds. He also devised a theory called *tagmemics.* This was a model for describing different languages using discrete elements (called slot/function and filler/class elements). Pike applied tagmemics to all levels of the grammar, from phonetics to discourse.

As a phonetician, Pike is known for his pioneering work modeling English intonation in a series of discrete levels. Pike also introduced the concepts of contour and register tones. Pike (1945) first used the terms stress-timed and syllable-timed. Perhaps most importantly, Pike formulated the distinction between *-emic* (as in phonemic) and *-etic* (as in phonetic). This has provided the modern conceptual difference between a phone (undifferentiated speech sound) and a phoneme (systematic unit of sound in a language).

Ken Pike's contributions to the field of linguistics combined with his dedication to the minority peoples of the world brought him numerous honors. In addition to his linguistic achievements, Pike was a devout Christian who contributed to Bible translation, poetry, and philosophy. He viewed his lifetime work as integrating his religious and linguistic passions, making him (in his own words!) "part horse, part donkey, a mule!"

An interesting issue for all these computations concerns the basic interval to be measured. Although vowel durations are a logical starting point, some researchers have suggested that other candidates should be considered. For example, some researchers, such as professors Francis Nolan and Eva Liina Asu, have explored the metrical foot (a basic timing measure).

Chapter 17

Coming from the Mouths of Babes

In This Chapter

- Tracking children's speech patterns
- Distinguishing healthy and disordered speech processes
- Applying this knowledge for transcription

Adults aren't the only people you'll transcribe in your phonetics classes and in your real-world career. For anyone working in speech language pathology, understanding child language is a must. The same holds true for anyone interested in the fields of childhood education, child language research, or dialectology. In this chapter, I take you through the periods of (healthy) speech development, discuss key differences between healthy and disordered speech, and give you some tips on how to put this knowledge into practice in your transcriptions.

Following the Stages of a Healthy Child's Speech Development

Knowing how children's speech develops is an important part of phonetics. Here you can track the sounds produced by children from the age of 6 months to 2 years old. I highlight universal aspects of young children's speech production and touch on some of the theories proposed to account for these amazing aspects of children's behavior.

Focusing on early sounds — 6 months

The first sounds to come out of a young infant are shaped by the physical capabilities of that very young person. When you're only a few months old, you don't have much of an adult-like vocal tract. The larynx is high in the throat and only begins to descend to adult-like proportions at approximately 5 to 7 months. At this stage in a person's life, these sounds are pretty much limited to high-pitched squeals, grunts, and cries.

Nevertheless, children at this age engage in a remarkable amount of communication, despite the inability to form words. They communicate with gaze, by imitating the pitch of their caretakers' speech, by making facial expressions, and by gesturing.

Babbling — 1 year

By approximately one year of age (often starting around 9 months), children begin the phase known as *babbling*, producing short, repeated utterances. This behavior, much beloved by parents, plays a major role in infant-parent bonding behavior.

Babbling is broadly described as having two phases:

- **Reduplicative:** This term refers to repeated speech. An example of reduplicative babbling would be "ba-ba-ba-ba" or "goo-goo-goo."
- **Variegated:** This term refers to many different sounds. Variegated babble consists of longer strings and more varied sounds than reduplicated babble. Some researchers also describe a jargon phase (occurring at about 10 months of age) at which adult-like stress and intonation begin to kick in. An example of variegated babbling would be "ka-be-to-gi-ta-ge."

Children babble when they're relaxed and comfortable. This behavior is thought to be a way of engaging the yet-developing vocal folds. Early babbling isn't necessarily related to communication, although babbling carries over into early word production.

Speech babble has provided researchers valuable insights into infant behavior. For instance, the rhythmic opening and closing gestures of children's mouths in forming utterances such as "buh-buh" and "ga-ga" have been interpreted in the *Frame-Content Theory.* This theory teases out the rhythmic opening and closing (syllabic) part of infant babbling behavior (called the *Frame*) from the segment-specific elements (such as consonants and vowels), called the *Content*. According to this view, a babbled syllable isn't a random mix of consonants and vowels, but instead motoric constraints result in the following pairs:

- Alveolar consonant and front vowel (such as /di/ and/de/)
- Labial consonant and central vowel (such as /bʌ/ and /ba/)
- Velar consonant and back vowel (such as /go/ and /gu/)

So far, researchers have found such patterns in English-speaking infants and in child speakers of other languages (including Swedish, Japanese, Quechua, Brazilian-Portuguese, Italian, and Serbian). These findings have spurred on

other researchers to investigate to what degree babbling is shaped by the growth of the vocal tract itself versus other developmental processes, such as the maturation of the motor control system (or the need for infants to first discover and then fine-tune relationships between their speech movements and sounds).

Researchers have also found that young children open the right side of their mouths more when they babble, suggesting that the left side of the brain controls this babbling.

Due to the physiological limits of young children, some sounds tend to be produced more than others. A study of 15 different languages, including English, Thai, Japanese, Arabic, Hindi, and Mayan, showed the following consonants commonly occur:

/p/, /b/, /m/, /t/, /d/, /n/, /s/, /h/, /w/, /j/

However, these phonemes were rarely found:

/f/, /v/, /θ/, /ð/, /ʃ/, /ʒ/, /ʧ/, /ʤ/, /l/, /ɹ/, /ŋ/

These data suggest that early babbling is at least partly independent of language-particular factors.

Forming early words — 18 months

Hearing a child's first words is one of the most rewarding experiences of being a parent. For a phonetician, studying the sound patterns in those first words is just about as exciting.

Young children can hear sound contrasts well before they can produce them. Just because they have immature articulatory systems doesn't mean that their sharp little minds aren't doing well at teasing out the sounds big people are telling them.

In terms of what children want to say, the most common items in the first 50 words are typically nouns, including such words as "daddy," "mommy," "juice," "milk," "dog," "duck," "car," "book," and "blocks." Young children follow with verbs and adjectives, including properties ("all gone," "more," and "dirty"), actions ("up," "down," "eat," "seat," and "go"), and personal-social terms ("hi," "bye," "please," and "thank you"). By the time children have acquired 50 words or so (usually by around 18 months of age), they start to adopt fairly regular patterns of pronunciation.

Although children vary a good deal in terms of the order in which they master speech sounds in production and perception, the following general tendencies seem to exist:

- As a group, vowels are generally acquired before consonants (by age three).
- Stops tend to be acquired before other consonants.
- In terms of place of articulation, labials are often acquired first, followed (with some variation) by alveolars, velars, and alveo-palatals. Interdentals (such as /θ/ and /ð/) are acquired last.
- New phonemic contrasts occur first in word-initial position. Thus, the /p/ to /b/ contrast, for instance, shows up in pairs such as "*p*at" and "*b*at" before "ca*p*" and "ca*b*."

Toddling and talking — 2 years

A two-year-old is a very different creature than a six-month old. The motoric and cognitive systems are much further developed (and, true, they generally relish saying "no!"). This section describes the sound inventory you can expect in English for a two-year-old talker.

By age 2, a typical English-speaking child has the following inventory of consonant phonemes:

- **Oral stops:** /p/, /t/, /k/, /b/, /d/, and /g/
- **Nasals:** /m/ and /n/
- **Fricatives:** /f/ and /s/
- **Approximants:** /w/

Still to be acquired are the interdental fricatives (/θ/ and /ð/) and the voiced alveo-palatal fricative (/ʒ/). These sounds are typically acquired after age 4.

In general, the relative order in which children acquire sounds reflects the sound's distribution in the world's languages. The sounds that are acquired early tend to be found in more languages, whereas the sounds that are acquired late tend to be less common across languages.

The first language: A king's experiment

Many people wonder what kind of language children might develop on their own. This kind of experiment (focusing on the *nature* part of *nature versus nurture*) was actually tried several times throughout history. The most famous case was Psamek I of Egypt (more commonly known as Psammetichus), a pharaoh living in the 7th century BCE. This story was described hundreds of years later in the writings of the Greek historian, Herodotus.

It goes like this: Psammetichus wanted to test whether the Egyptians were the most primitive race, so he came up with an experiment. He took two children, born from commoners, and had a shepherd raise them. The shepherd was strictly forbidden to allow any spoken word to be said in their presence. The goal was to see what the first word uttered by the children would be, assuming they had no other model to pattern after.

According to Herodotus, when the children were brought before Psammetichus, one of them said something that sounded like *bekos*, the Phrygian word for bread. From this, Psammetichus concluded that the capacity for speech is innate and the natural language of people was Phrygian.

If this account is true, phoneticians should praise the king for at least two things: First, his intellectual honesty. Instead of somehow determining that the Egyptians were the first people (which he likely hypothesized), he was instead informed by the data and concluded in favor of the Phrygians. Second, the king concluded that speech is innate — a view that continues to be influential today.

Today, scientists know that infants come into the world with the capacity to acquire any language. Thus, a German baby from Ulm could be dropped in to parents in Kenya and quickly be speaking Kikuyu (or vice versa). No one inborn tongue exists.

Knowing What to Expect

Everyone knows that, compared to adults, children make mistakes in their speech. However, determining whether a child's speech is healthy or disordered isn't as easy. Because children acquire speech structures over time, certain errors are expected at certain ages. These normal (healthy) patterns of development can be contrasted with disordered child language processes.

A basic way to start thinking about whether a child's speech is disordered (and a question familiar to many parents) is to ask: What sounds should my child be saying at such-and-such age? When answering this question, clinicians consider children's phonological processes when evaluating healthy and disordered patterns of development, which I explain in the following sections.

Eyeing the common phonological errors

Phonologists begin by studying the errors that healthy children make when learning language. These data show many commonalities across languages, including languages from very different language families. Phoneticians generally agree that children's phonological errors include the following:

- Boo-boos at the level of syllable production
- Substitutions of one consonant or vowel segment for another of like kind
- *Assimilation* processes, in which one sound becomes more like one another

Table 17-1 gives you some examples:

Table 17-1 Common Childhood Errors

Syllable-Level Processes	***Example***	***Production (IPA)***
Weak syllable deletion	"potato"	/ˈtedo/
Final consonant deletion	"book"	/bʊ/
Reduplication	"baby"	/bibi/
Cluster reduction	"climb"	/kaɪm/
Substitutions		
Stopping	"soup"	/tup/
Fronting	"cake"	/tek/
Deaffrication	"jump"	/ʒʌmp/
Liquid gliding	"like"	/waɪk/
Vocalization (liquid becomes vowel)	"line"	/jajn/
Assimilatory Processes		
Labial	"pot"	/pɑp/
Alveolar	"mine"	/naɪn/
Velar	"harden"	/ˈhargn̩/
Prevocalic voicing	"tap"	/dæp/
Devoicing	"ride"	/ɹaɪt/

This table contains examples that probably seem familiar or even cute to the average person. For example, saying /ˈtedo/ for "potato." An adult may create these kinds of errors when trying to imitate child speech.

Examining patterns more typical of children with phonological disorders

Child language specialists also seek to determine patterns that can serve as a warning of phonological disorders in children. Experts differ somewhat on the best ways to classify these disorders; however, they generally agree on the types of underlying problems. Two key concepts include

- Certain children may have a language *delay* by showing persisting normal processes that last longer than they are supposed to.
- Some children show unusual, idiosyncratic, or atypical *deviance* in the application of phonological rules, compared to other children.

Table 17-2 shows some examples of idiosyncratic phonological processes in child language:

Table 17-2 Errors Suggesting Childhood Speech Disorders

Disorder	*Example*	*Production (IPA)*
Glottal replacement	"stick"	/stɪʔ/
	"better"	/ˈbɛʔɛ/
Backing	"test"	/kɛst/
	"smash"	/smæg/
Initial consonant deletion	"guess"	/ɛs/
	"kiss"	/ɪs/
Stops replacing a glide	"yellow"	/ˈdɛdo/
	"wait"	/bet/
Fricatives replacing a stop	"quit "	/kwɪs/
	"duck"	/zʌk/

These idiosyncratic cases wouldn't likely be included in the average adult's imitation of child speech. The typical parent probably wouldn't always know what is normal and what is worrisome, hence why he or she should seek a professional opinion.

Children with recognizable speech errors may have the following disorders:

- **Speech sound disorders:** These disorders include both articulatory errors and problems with phonological development.
- **Childhood apraxia of speech:** A motor speech disorder in which children know what they want to say but have difficulty mapping these intended sounds into realized speech movements.
- **Dysarthria:** A motor speech disorder involving problems with the muscles of the mouth, face, or respiratory system.
- **Orofacial myofunctional disorders:** Also known as *tongue thrust,* these disorders involve an exaggerated protrusion of the tongue during speech and/or swallowing.
- **Stuttering:** A fluency problem marked by disruptions in the production of speech sounds that can impede communication.
- **Voice disorders:** They include problems in producing sound at the level of the larynx.

For more information about these different disorders, contact the following organizations:

- `www.asha.org/public/speech/disorders/childsandl.htm` (United States)
- `www.caslpa.ca/` (Canada)
- `www.rcslt.org/` (United Kingdom)
- `www.asha.org/members/international/intl_assoc.htm` (Other countries, from Argentina to Vietnam)

Transcribing Infants and Children: Tips of the Trade

The exact reasons *why* you're transcribing can guide you in the tools to use and in the way you do your transcription. If you're creating transcriptions (from recordings) for clinical or teaching purposes, then you have many possible options to choose from. For example, you can be more or less *narrow* (transcribing fine-grained detail), incorporate certain characters from the *ExtIPA* (extensions of the IPA), and use a variety of different conventions to represent *prosody* (melody) — (see Chapters 10 and 11 for more information).

However, if you're working in a lab or clinic that has an established protocol, you need to master those specific tools. In this section, I introduce you to a variety of methods and techniques that can be useful. I also provide you some brief examples to get you started. I include speech from the period of early word acquisition (9 to 16 months). In addition to these examples of healthy speech, I also provide a snippet of speech from a 2-year-old child with a cochlear implant to show how speech presents as children adapt to prosthetic hearing.

Delving into diacritics

In a perfect world, cleanly articulating children would produce only lovely substitution errors for your corpus. You would then transcribe little Jimmy's production of /fɪs/ for *fish*, consider it a backing error (see Table 17-1), and feel darn good about yourself.

However children's actual speech is far messier. There are errors both at the phonemic (such as substitutions, or *metathesis*, the switching of sounds) and phonetic (for instance distortions and coarticulatory) levels. You typically need to complete a s*ystematic narrow transcription*, indicating allophonic variation of individual phonemes. This usually requires the use of several *diacritics*, marks to fine-tune transcription. I introduce diacritics in Chapter 3 and further describe them in Chapter 19.

Table 17-3 lists diacritics useful for working with children's speech, sorted by voicing, place, and manner of articulation.

Table 17-3 Diacritics for Transcribing Children's Speech

Feature	*Diacritic*	*Symbol*	*Example*		*Applications and Hints*
Voicing	Partially voiced	[◌̬]	"sister"	[ˈsɪs̬tɚ]	Not as common as devoicing instances.
	Partially devoiced	[◌̥]	"pin" "sick"	[m̥ĩn] [s̥ɪʔk]	Seen in cleft-palate speech substitutions.
Place	Advanced tongue position	[◌̟] or [◌˖]	"hot"	[hɑ̟t]	Best to refer to a vowel quadrilateral when working with these symbols. Regional dialects also greatly affect these changes in vowel quality.
	Retracted tongue position	[◌̠] or [◌˗]	"hat"	[hæ̠t]	
	Raised tongue position	[◌̝]	"bet"	[bɛ̝t]	
	Lowered tongue position	[◌̞]	"look"	[lʊ̞k]	
Secondary Articulations	More rounded	[◌̹]	"seat"	[si̹t] (or [syt]**)	**Some phoneticians use the rounded phoneme symbols ([y]) instead.
	Less rounded	[◌̜]	"boot"	[bu̜t]	
	Nonlabialized (lip spread)	[◌̟]	"ship"	[ʃ̟ɪp]	Seen in some children with phonological disorders.
	Dentalized	[◌̪]	"rich" "jam" "zebra"	[ɹɪ̪θ] [ðd̪æ̃m] [ˈz̪ibɹə]	Seen in some children with phonological disorders. Frontal lisp.
	Palatalized	[ʲ]	"sick"	[sʲɪk]	Frequent distortion of [s] seen in children.
	Lateralized	[ɬ] [ɮ]	"sip" "zip"	[ɬɪp] [ɮɪp]	Lateral lisp.
Stop Release Features	Unreleased stops	[˺]	"bat"	[bæt˺]	[˺] is common in GAE for consonants in syllable-final position and [ʰ] for consonants in syllable-initial position. Thus, only note if pervasive and unusual. Best to note *during* the recording session.
	Aspirated stops	[ʰ]	"Pam"	[pʰæm]	
	Unaspirated stops	[˭]	"pig"	[p˭ɪg]	Use if a child has a pervasive lack of aspiration for voiceless stops.
Manner	Nasal emission	[◌̃ʰ]	"snake"	[sʰ̃nek]	Sound of air escaping through nose when it shouldn't.
	Denasalized	[◌̃˟]	"my new song"	[m̃ai ñu s̃ɔ̃ŋ]	Sounds like having a stuffy nose; air can't escape the nose when it should.

Here are some more practical tips for when transcribing children's speech:

- **Don't become frustrated.** You can't be expected to identify every phoneme your talker produces.
- **Circle the features you do know, work on the rest later.** For instance, if you know the phoneme is a voiced fricative, you can write:

- **Take frequent breaks.** Don't listen to a sound more than three times in a row.
- **Keep your mind clear and don't read into the transcription what is not there.** I have seen many transcriptions that reflect what the transcriber thought (or desperately hoped) would come next.

Turning to technology for transcription

In addition to the individual phonetician working with an IPA chart, a dizzying array of computer-based programs designed to help researchers is available. A partial list includes Alembic, CHILDES, DAISY, Digital Lava, Discourse Transcription, Emu, Festival, GATE, Hyperlex, Informedia, ISIP, LDC, LIPP, MacSHAPA, MediaTagger, ODF, Praat, SALT, SGREP, Tipster, TreeBank, VoiceWalker, and UTF.

In an effort to systemize these data, Professor Brian MacWhinney of Carnegie Mellon University is coordinating a project called *TalkBank*, designed to create shared databases of primary materials between several disciplines that study human communication. Check out `http://talkbank.org/` for more information.

In the meantime, you may find yourself working on a project, transcribing using a set format within a given program. Many of these programs have IPA-compatible modules that allow phonetic labeling for various purposes. For instance, *Praat* is a toolkit for phonetic analysis developed by Paul Boersma and David Weenink at the University of Amsterdam. Praat allows the user to insert IPA characters for a variety of purposes, including the labeling of spectrograms and intonation plots. Look up Praat at `www.fon.hum.uva.nl/praat/`.

Other programs, such as *The Logical International Phonetics Program* (LIPP, Intelligent Hearing Systems), are designed for phonetic transcription and analysis. LIPP allows users to type in IPA characters by a variety of pull-down menus or with a keyboard that lets users change and customize the characters displayed on each key. LIPP also contains a phonetic dictionary, where you can type in the letters of a word, such as "cow," and have the system return a phonetic transcription, such as /kaʊ/. Sound classifications may then be completed in a variety of formats. You can find screenshots of the transcription process at `www.ihsys.com/brochures/brochure_lipp.pdf`.

Study No. 1: Transcribing a child's beginning words

The first sample is from a project performed by professor Marilyn Vihman at Stanford University (currently at the University of York, England), investigating the beginning of children's phonological organization. Table 17-4 shows transcriptions that come from a young child babbling as she approached her first words.

Table 17-4 Transcriptions from a Young Child's Speech

Age	*Vocabulary Stage*		*Examples*	
		"baby"	"elephant"	"blanket"
9 to10 months old	Beginning of lexical use.	[pɐpɛː], [əp^{h}æːp(^{h})æ], [tɛɪtɪː],[te^{ɪ}tɛ]	--	--
15 months old	Some organization can be determined.	[be bi], [(hə)bebi]	[ʔæɪnu], [ʔaijɐ̃]	[baji],[hnba]
16 months old	Rapid lexical advance begins.	[(ə)beɪbi]	[ʔaiː(n)jʌ], [eɐ̃jɐ̃]	[kɛt]?

These transcriptions include parentheses for sounds produced quietly (hə) and (ə), and light aspiration is shown with a superscript "h" in parentheses. Vowel lengthening (using the diacritic [ː]), glottal stop, and nasalization are noted. A question mark after "[kɛt]" indicates the transcriber was unsure of this transcription.

Study No. 2: A child with a cochlear implant (CI)

The second study performed by Andrea Warner-Czyz, PhD, at the University of Texas at Dallas, includes data from a young girl, *H*, profoundly deaf from birth, fitted with a cochlear implant (CI), activated when she was 11 months 22 days old. This girl was considered a successful CI user. The following minitable shows some utterances transcribed 13 and 18 months post-implant.

Time Post Implant	*Parent*	*Child Response (IPA)*
13 months post	Mommy, see the baby.	/ma mi ʃi ə be bi/
18 month post	Pick him up	/i jə bəp/
18 month post	Hey, Mommy. Sit down.	/eː mami di do/

The key purpose of these data was to identify basic errors (at the phonemic level) and to track the expansion of the child's phoneme repertoire. As such, the researcher conducted a fairly broad transcription. Features such as vowel length were detailed, using [ː] for long vowels and elsewhere [ːː] for extra-long vowels. Patterns of omission/substitution/metathesis were described, and unexpected patterns of intonation are indicated. In most cases, phonetic departure from targets is indicated with substituted IPA symbols (for instance, /ʃ/ found for the /s/ target of "see" in the line 1).

Babbling birds: Clues to speech mysteries

Birds babble. Not just in cartoon commercials, but they also babble when young songbirds are acquiring their songs. It depends on the exact species and how they're exposed to their adult songs. Here scientists start to see possible connections to human behavior.

The Zebra finch is a particularly well-studied species. Young male finches first have an auditory learning period in which they must hear examples of the songs they're to sing. They then produce a variety of immature songs called *subsong.* The birds then advance to the plastic stage (where some of the adult forms are noticeable), followed by mature song.

Subsong seems to be like human babble in several ways:

- For both babies and birds, youngsters are attracted to their own species more than to others. Human babies are more interested in human voices than other noises, and baby birds are far more tuned in to the songs of their own species than the songs of other types of birds.
- These phases of birdsong learning seem parallel to humans having a sensitive period for language learning (when language learning is best accomplished, after which point it becomes more difficult and learners are left with a non-native accent).
- Like humans, if song patterns are reinforced with positive social feedback, they're more likely to recur, which is especially prominent in social species such as the Zebra finch.
- Auditory feedback plays an important role in maintaining both speech and birdsong, although researchers don't fully understand the processes involved.

Scientists are exploring other interesting commonalities between bird communication and human speech, including similar neural bases and genetic contributions (including the FOXP2 gene). Birdsong has syllables, dialects, and accents. Scientists can learn a lot by studying feathered friends. Birdbrain isn't necessarily an insult.

To hear song sparrows learn to sing, go to `http://birdnote.org/show/song-sparrows-learn-sing`.

Chapter 18

Accentuating Accents

In This Chapter

- Defining dialectology
- Mapping English accents in the United States
- Getting a sense of other world Englishes

A world without speech accents would be flat-out dull and boring. Actors and actresses would lose their pizzazz, and people would have nobody to tease for sounding funny. All joking aside, accents are extremely interesting and fun to study because, believe it or not, everyone has an accent. Understanding accents helps phoneticians recognize the (sometimes subtle) differences speakers have in their language use, even when they speak the same language.

This chapter introduces you to the world of dialectology and English accents. You peer into the mindset of a typical dialectologist (if such a thing exists) to observe how varieties of English differ by words and by sounds. You then hop on board for a whirlwind tour of world English accents. Take notes and you can emerge a much better transcriber. You may even pick up some interesting expressions along the way.

Viewing Dialectology

People have strong feelings concerning different accents. They tend to think that their speech is normal, but other folks' speech sounds weird. This line of thinking can go the other extreme with people thinking that they have a strong country or city accent and that they won't ever sound normal.

Think of the times you may have spoken to someone on the phone and reacted more to the way they sounded than based on what the person actually said. Awareness of a *dialectal difference* is still a strong feeling many people have. In fact, some phoneticians may argue that judging people based on their dialect is one of the few remaining socially accepted prejudices. Although most people have given up judging others based on their ethnic

background, race, gender, sexual orientation, and so forth (at least in public), some people still judge based on dialect. Along comes a *Y'all*!, *Oi!,* or *Yer!* and there is either a feeling of instant bonding or, perhaps, repulsion.

To shed some light on this touchy subject, *dialectologists* study differences in language. The word *dialect* comes from the Greek *dia-* (*through*) and *-lect* (*speaking*). To dialectologists, a language has regional or social varieties of speech (classified as a *lect*). For example, the United States and Britain have noticeable differences between speaking styles in the South and North (geographic factors). Social speech differences, such as what you may find comparing a tow truck driver and a corporate attorney, also exist.

Furthermore, a village or city may have its own lect. According to this classification system, each individual has his or her own *idiolect*. Note, an idiolect *isn't* the speech patterns of an idiot (although, I suppose an idiot would have his or her own idiolect, too).

Mapping Regional Vocabulary Differences

Dialectologists create dialect maps showing broad dialect regions, such as the West, the South, the Northeast, and the Midwest of the United States. Within these broad areas, they create further divisions called *isoglosses*, which are boundaries between places that differ in a particular dialect feature.

Mapping pronunciation differences: Greasy or greazy?

Dialects can get picky, such as in how people in different regions of a country pronounce particular words. For example, take the word "greasy." The way people of different regions pronounce this word marks a clear boundary between major dialect regions of the United States. In the North and West, "greasy" is pronounced with [s]. In the South and Midland regions, it's pronounced with [z].

According to the Dictionary of American Regional English, the "greasy" region extends from the deep South to southern parts of New Jersey, Pennsylvania, Ohio, Indiana, and Illinois and all of Missouri, Texas, and New Mexico. The verb "grease" also follows this pattern. However, the noun "grease" is uniformly pronounced with [s].

Other isogloss bands mark different regions. For instance, a "pin/pen" band sweeps down below the Mason-Dixon line in the United States, incorporating West Virginia, Kentucky, Tennessee, and much of Arkansas, Oklahoma, and Texas. Within this region, talkers merge /ɪ/ and /ɛ/ before /n/, pronouncing both phonetically as [ĩn]. This kind of sound change is called a *merger*, in that sound changes are neutralized. Such sound patterns, most common in North Carolina, appear to be a relic of 17th century colonial English.

Lexical (vocabulary) *variation* plays an important role in dialectal differences. Such variability is common throughout the languages of the world. To study these variations, dialectologists create regional dialect maps at the lexical level by collecting samples of the way people name certain objects (such as common people, places, and things). A group of people has the same dialect if they share many of the words for things. For instance, people use different words for "dragonfly" along the American Atlantic coast, illustrating isogloss boundaries between words such as "darning needle," "mosquito hawk," "spindle," "snake feeder," "snake doctor," and "snake waiter." Actually, you may even call them "eye-stitchers" (in Wisconsin), "globe-skimmers" (in Hawaii), or "ear sewers."

You can test how you weigh in on this kind of vocabulary variation with this question designed for North Americans:

> What do you call a large, made-to-order sandwich on a 6-inch roll?
>
> a.) Hero
>
> b.) Hoagie
>
> c.) Po-boy
>
> d.) Sub
>
> e.) Other

Your answer likely depends on where you live and on your age. If you're from New York City, you may answer "hero." If you're from Philadelphia, you may answer "hoagie." If you're from Texas or Louisiana, you may answer "po-boy." The usual champ, "sub," now seems to be edging out the other competitors, especially for younger folks.

If you answer other, you may refer to this sandwich by a wide variety of names, such as "spucky," "zep," "torp," "torpedo," "bomber," "sarney," "baguette," and so on. For color maps of how approximately 11,000 people responded to this type of question, check out `www4.uwm.edu/FLL/linguistics/dialect/staticmaps/q_64.html`.

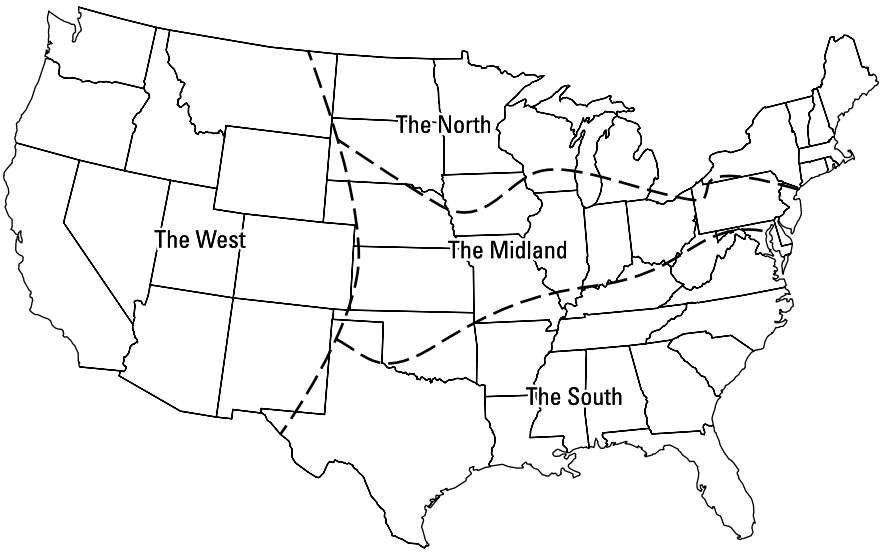
The Dictionary of American Regional English, a federally funded project by the University of Wisconsin, provides excellent up-to-date information about lexical variability in American English. This group has published a five-volume dictionary and maintains a website with sound samples, educational materials, and online quizzes. Take its vocabulary quiz at `http://dare.news.wisc.edu/quiz/`. You can find other similar sites for a few other varieties of English at the following:

- **Australian:** `www.abc.net.au/wordmap/`
- **British:** `www.bbc.co.uk/voices/`
- **Canadian:** `http://dialect.topography.chass.utoronto.ca/dt_orientation.php`

Transcribing North American

Dialectologists differ when it comes to dividing up the United States into distinct regional dialect areas. Some favor very broad divisions, with as little as two or three regions, while others suggest fine-grained maps with hundreds of regional dialect areas.

I follow the divisions outlined in the recently completed Atlas of North American English, based on the work of dialectologist William Labov and colleagues. This atlas is part of ongoing research at the University of Pennsylvania Telsur (telephone survey) project. The results, which reflect more than four decades of phonetic transcriptions and acoustic analyses, indicate four main regions: the West, the North, the South, and the Midland. Figure 18-1 shows these four regions.



Map by Wiley, Composition Services Graphics

Figure 18-1: The United States divided into four distinct regional dialect areas.

The first three regions have undergone relatively stable sound shifts, whereas the Midland region seems to be a mix of more variable accents. The following sections look closer at these four regions and the sound changes and patterns that occur in the speech of their locals.

The West Coast: Dude, where's my ride?

The area marked West ranges from Idaho, Wyoming, Colorado, and New Mexico to the Pacific coast. This large region is known mostly for the merger

of /ɑ/ and /ɔ/ (for example "cot" versus "caught" and "Don" versus "Dawn"), although this blend is also widespread in the Midland. A common feature of the West is also fronting of /u/. For example, Southern Californian talkers' spectrograms of /u/-containing words, such as "n*ew,*" show second formants beginning at higher-than-normal frequencies (much closer to values for /i/).

In general, these characteristics mark the West:

- **Rhotic:** *Rhotic* dialects are ones in which final "r" sound consonants are pronounced. For instance, the "r" in "butte*r.*"
- **General American English (GAE):** This is perceived to be the standard American English accent. It's typically the accent you would hear used by news anchors.
- **Dialectal variability mainly through stylistic and ethnic innovations:** Most of the variation in dialect is due to social meaning (style) or variants used by different ethnic groups in the area.

A rather stereotyped example of such variation is the California surfer, a creature known for fronting mid vowels such as "b*u*t" and "wh*a*t," pronouncing them as /bɛt/ and /wɛt/. Expressions such as "I'm like . . ." and "I'm all . . ." are noted as coming from young people in Southern California (the *Valley Girl* phenomenon). Linguists describe these two particular creations as *the quotative,* because they introduce quoted or reported material in spoken speech.

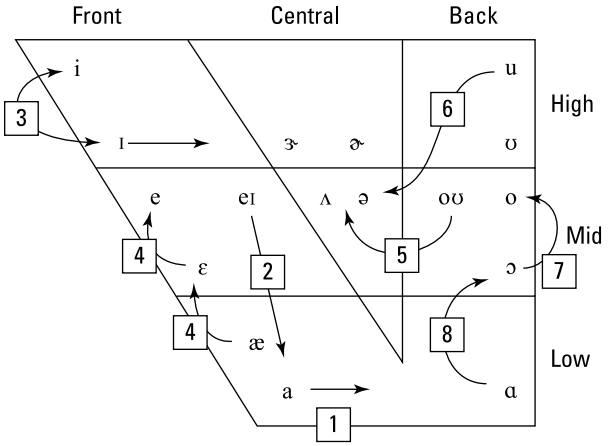
Other regionalisms in the West may be attributed to ethnic and linguistic influences, for example the substitution of /ɛ/ to /æ/ (such as "elevator" pronounced /ˈælɪvedɚ/) among some speakers of Hispanic descent, and more syllable-based timing among speakers from Japanese-American communities.

The South: Fixin' to take y'all's car

The Southern states range from Texas to Virginia, Delaware, and Maryland. This accent has striking grammatical ("fixin' to" and "y'all") and vocabulary characteristics ("po-boy").

In general, these characteristics mark the South:

- **Rhotic:** However, some dialects of Southern states' English are more non-rhotic.
- **Lexically rich:** This dialect has a plentiful, unique vocabulary.
- **Vowels:** One of the most distinct qualities of Southern American English is the difference in vowels compared to GAE. An important phonetic feature of the Southern accent is the Southern vowel shift, referring to a *chain shift* of sounds that is a fandango throughout the vowel quadrilateral. Figure 18-2 shows this chain shift.



Map by Wiley, Composition Services Graphics

Figure 18-2: Southern vowel shift

Follow these steps and see if you can make this vowel shift:

1. **Delete your [**aɪ**] diphthong and substitute an [**ɑ**] monophthong.**

   "Nice" becomes [nɑs].

2. **Drop your [**eɪ**] tense vowel to an [**aɪ**].**

   "Great" becomes [gɹaɪt].

3. **Merge your [**i**]s and [**ɪ**]s before a nasal stop.**

   "Greet him" now is [gɹɪt hĩm].

4. **Merge your [**æ**]s and [**ɛ**]s.**

   "Tap your step" becomes [tʰɛp jəˠ steɪp].

5. **Swing your [**æ**] all the way up to [**e**].**

   "I can't" becomes [aɪ kẽnt].

6. **Move your back vowels [**u**]s and [**o**]s toward the center of your mouth.**

   "You got it" becomes [jə ˈgʌt ɪt].

7. **Raise the [**ɔ**] up to [**o**] before [**ɹ**].**

   "Sure thing" becomes [ʃoəˠ θaĩŋ].

8. **Raise [**ɑ**] to [**ɔ**] before [**ɹ**].**

   "It ain't hard" becomes [ɪʔ eĩnˈt hɔəˠɹd].

   Congratulations.

   [weɬ ðə ˈmaĩn θaĩŋɪz| jə ˈspɪkĩn ˌsʌðə̃n‖]

"Well the main thang is ya speakin' southen" (which means "Well the main thing is you're speaking Southern," written in a Southern accent).

In old-fashioned varieties of Southern states English (along with New England English and African-American English), the consonant /ɹ/ isn't pronounced. Think of the accents in the movie *Gone with the Wind*. Rather than pronouncing /ɹ/, insert a glided vowel as such:

"fear" as [fiə]

"bored" as [boəd]

"sore" as "saw" [soə]

Another Southern states' consonant feature is the /z/ to /d/ shift in contractions. The voiced alveolar fricative (/z/) is pronounced as a voiced alveolar stop (/d/) before a nasal consonant (/n/). In other words:

"isn't" as [ˈɪdn̩t]

"wasn't" as [ˈwʌdn̩t]

Are Texans losing their twang?

Dialects change. If you watch an old black-and-white movie, you may notice that the accents and expressions sometimes clash with the way people talk today. But how much has been lost, where, and by whom? Getting these answers is the job of dialectologists. Lars Hinrichs (an associate professor at The University of Texas at Austin) is systematically comparing the speech of University of Texas students with a database collected 30 years ago by the late Professor Gary Underwood, a professor of English linguistics. Data have been collected on vocabulary (words like "lightning bug" for "firefly," still in use) and accent (including the "pit/pet" and "cot/caught" mergers (for [ɪ]/[ɛ] and [ɑ]/[ɔ]).

Altogether, recordings of more than 700 speakers of Texas English have been collected, half of them in the early 1980s and the other half around the year 2012. The recordings allow Hinrichs to track the recent change of Texas English over the past 30 years.

So far, the data indicate that GAE has infiltrated Texas dialects. As usual, young women, the segment of the population typically at the vanguard of accent revision (refer to Chapter 15), are largely heading this change.

Young women are the first to adopt GAE prestige forms, preferring them over old, Texan-sounding forms, and then pass them on to other speakers, which is how Texas English changed from a mostly non-rhotic to a rhotic (r-colored) variety, and it's how dialects are changing everywhere.

Texas English is quite different from other American dialects in other ways as well. Middle-class speakers apparently have adopted the new GAE forms, whereas both working class and upper class Texans (the poorest and the richest speakers) tend to hold on to their Texas twang longer. Texan-sounding speech may be a (fairly unusual) case of a vernacular dialect that enjoys high social prestige. For illustration, think of high-status individuals such as President George W. Bush or Texas governor Rick Perry, both of whom speak in unmistakably local accents of Texas English.

For more information, see The Texas English Project at `www.texasenglish.org`.

The South is teeming with characteristics that dialectologist enjoy arguing over. Some dialectologists classify different varieties of Southern states English including Upper South, Lower South, and Delta South. Others suggest Virginia Piedmont and Southeastern Louisianan. Yet others disagree with the classifications of the preceding varieties. Say what you will about the South, it's not boring linguistically.

The Northeast: Yinzers and Swamp Yankees

The Northeast region has a wide variety of accents, strongest in its urban centers: Boston, New York, Philadelphia, Buffalo, Cleveland, Toledo, Detroit, Flint, Gary, Syracuse, Rochester, Chicago, and Rockford. Dialectologists identify many sub-varieties, including boroughs of New York City.

Some key characteristics for this region include the following:

- **Derhoticization:** The loss of r-coloring in vowels. This is especially the case in traditional urban areas like the Lower East Side of New York City or in South Boston, whose English is non-rhotic.
- **Vowels:** Key differences include the Northern cities' vowel shift and the low-back distinction between [ɑ] and [ɔ].
- **Vocabulary distinctions and syntactic forms:** For example, *swamp Yankees* (hardcore country types from southern Rhode Island), and syntactic forms (such as "yinz" or "yunz" meaning "you (plural)," or "y'all" in Southern states accent).

The accent change in this region goes in the opposite direction than the accent in the Southern states (refer to previous section). It's a classic chain shift that begins with [æ] swinging up to [i], and ends with [ɪ] and [ɛ] moving to where [ʌ] was. Figure 18-3 shows the Northern cities shift. Follow these steps and pronounce all the IPA examples to speak Northeast like a champ.

1. **Change low vowel [æ] to an [iə].**

   "I'm glad" becomes [ə̃m gliəd].

2. **Move the back vowel [ɑ] to [æ].**

   "Stop that" becomes [stæp dæt].

3. **Move the [ɔ] to where [ɑ] was.**

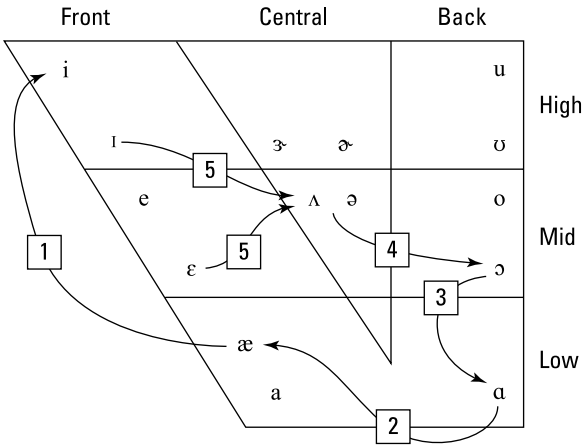
   "Ah, get out" becomes [ɑ: gɪt ɑt].

4. **Move central [ʌ] to where [ɔ] was.**

   "Love it" becomes [lɔv ɪt].

5. **Move the front [ɛ] and [ɪ] to center [ʌ]/[ə].**

   "Let's move it" becomes [ləts ˈmʊv ət].



Map by Wiley, Composition Services Graphics

Figure 18-3: Northern cities shift.

The Midlands: Nobody home

The Americans in the Midlands decline from participating in the Southern states' and Northern cities' craziness. In general, this dialect is rhotic. After that, life gets sketchy and difficult in trying to characterize this region.

The folks in this region are somewhat like the Swiss in Europe, not quite sure when or where they should ever commit. The dialect does exhibit some interaction between [i] and [ɪ] and between [e] and [ɛ], but only in one direction (with the tense vowels laxing). Thus the word "Steelers" is [ˈstɪlə˞z] and the word "babe" is [bɛb]. However, like the North, the diphthong [aɪ] is left alone. Thus, "fire" is mostly pronounced [faɪɹ], not [fɑɹ].

Perhaps seeking something exciting, some dialectologists have divided the midlands into a North and a South, with the North beginning north of the Ohio River valley. Dialectologists argue that the North Midlands dialect is the one closest to GAE, or the Standard American Accent heard on the nightly news and taught in school. In this region, the /ɑ/ and /ɔ/ (back vowel) merger is in transition.

The South Midlands accent has fronting of [o] (as in "road" [ɹʌd]). The accent also has some smoothing of the diphthong /aɪ/ toward /ɑ:/. As such, dialectologists consider South Midland a buffer zone with the Southern states.

Pittsburgh has its own dialect, based historically in Western Pennsylvania (North Midland), but possessing a unique feature: the diphthong /aʊ/ *monophthongizes* (or becomes a singular vowel) to /ɑ/, thus letting you go "downtown" ([dɑ̃nˈtɑ̃n]). St. Louis also has some quirky accent features, including uncommon back vowel features, such as "wash" pronounced [wɑɹʃ] and "forty-four" as [ˈfɑɹfɪ.fɑɹ] by some speakers.

Black English (AAVE)

Dialectologists still seem to be struggling for the best name for the variety of English spoken by some black Americans. Many linguists debate the appropriate term to classify this variant. Terms include Black English (BE), Black English Vernacular (BEV), African-American Vernacular English (AAVE), Ebonics (although highly out of favor), or Inner City English (ICE). Also called *jive* by some of the regular public, it's up for debate whether this dialect arose from a *pidgin* (common tongue among people speaking different languages), is simply a variety of Southern states English, or is a hybrid of Southern states English and West African language sources.

I go with *AAVE*. This variety serves as an *ethnolect* and *socialect,* reflecting ethnic and social bonds. Linguists note distinctive vocabulary terms and syntactic usage in AAVE (such as "be," as in "They be goin'" and loss of final "s," as in "She go").

Speakers of AAVE share pronunciation features with dialects spoken in the American South, including the following:

- **De-rhoticization:** R-coloring is lost.
- **Phonological processes:** For example, /aɪ/ becomes [aː]) and /z/ becomes [d] in contractions (such as "*is*n't" [ˈɪdn̩t]).
- **Consonant cluster reduction via dropping final stop consonants, with lengthening:** Examples include words, such as "ri*sk*" ([ɹɪːs]) and "pa*st*" [pæːs]), and words with (-ed) endings, such as "wa*lked*" [wɑːk].
- **Pronunciation of GAE /θ/ as [t] and [f], and /ð/ as [d] and [v]:** At the beginning of words, /θ/ becomes [t], otherwise as [f]. Thus, "a thin bath" becomes [ə tʰĩn bæf]. Similarly, /ð/ becomes [d] at the beginning of a word and [v], elsewhere, which makes "the brother" [də ˈbɹʌvə].
- **Deletion of final nasal consonant, replaced by nasal vowel:** The word "van" becomes [væ̃].

- **Coarticulated glottal stop with devoiced final stop:** The word "glad" becomes [glæːtʔ].
- **Stress shift from final to initial syllable:** The word "police" becomes [ˈpʰoʊlis] or [ˈpʰoʊ.lis].
- **Glottalization of /d/ and /t/:** The words "you didn't" become [ju ˈdɪʔn̩].

Canadian: Vowel raising and cross-border shopping

In terms of sound, Canadian English shares many features of GAE, including syllable-final rhotics (for example, "car" is [kʰɑɹ]) and alveolar flaps, [ɾ], as in "Betty" ([ˈbɛɾɪ]). Notable features not common in American English include the following:

- **Canadian raising:** *Canadian raising* is a well-studied trait in which the diphthongs /aɪ/ and /aʊ/ shift in the voiceless environment. For both of them, the diphthong starts higher. Instead of beginning at /a/, it begins at /ʌ/. Moreover, it typically takes place before voiceless consonants. Thus, these words (with voiced final consonants) are pronounced like GAE:
  - "five" as [faɪv]
  - "loud" as [laʊd]

  Whereas the following words get their diphthongs raised, Canadian style:
  - "fife" as [fʌɪf]
  - "lout" as [lʌʊt]
- **The behavior of /o/ and /ɛ/ before rhotics:** Canadian maintains the /o/ before /ɹ/, where a GAE speaker wouldn't. For "sorry," a GAE speaker would likely say [ˈsɑɹi], whereas a Canadian English speaker would say [ˈsoɹi]. You can listen to a Canadian produce these sounds at `www.ic.arizona.edu/~lsp/Canadian/words/sorry.html`.

  Although many Northeastern speakers in the United States distinguish /ɛ/ and /æ/ before /ɹ/ (such as pronouncing "Mary" and "merry" as [ˈmæɹi] and [ˈˈmɛɹi]), many Canadians (and Americans) merge these sounds, with the two words using an /ɛ/ vowel.

A good test phrase for general Canadian English:

"Sorry to marry the wife about now" [ˌsoɹi tə ˈmɛɹi ðə wʌɪf əˌbʌʊtˈ naʊ]

However, this phrase wouldn't quite work for all Canadian accents, such as Newfoundland and Labrador, because they're quite different than most in Canada, having more English, Irish, and Scottish influence. These dialects lack Canadian raising and merge the diphthongs /aɪ/ and /ɔɪ/ to [ɑɪ] (as in "line" and "loin" being pronounced [lɑĩn]). They also have many vocabulary and syntactic differences.

If all else fails, a phonetician can always fall back on the /æ/-split in certain loanwords that have [ɑ] in GAE. To see if somebody is from Canada, ask him or her how to pronounce "taco," "pasta," or "llama." If he or she has an /æ/ in these words, the person is probably Canadian.

Transcribing English of the United Kingdom and Ireland

Describing the English dialects of the United Kingdom and Ireland is a tricky business. In fact, there are enough ways of talking in the British Isles and Ireland to keep an army of phoneticians employed for a lifetime, so just remember that there is no one English/Irish/Welsh/Scottish accent. This section provides an overview to some well-known regional dialects in the area.

England: Looking closer at Estuary

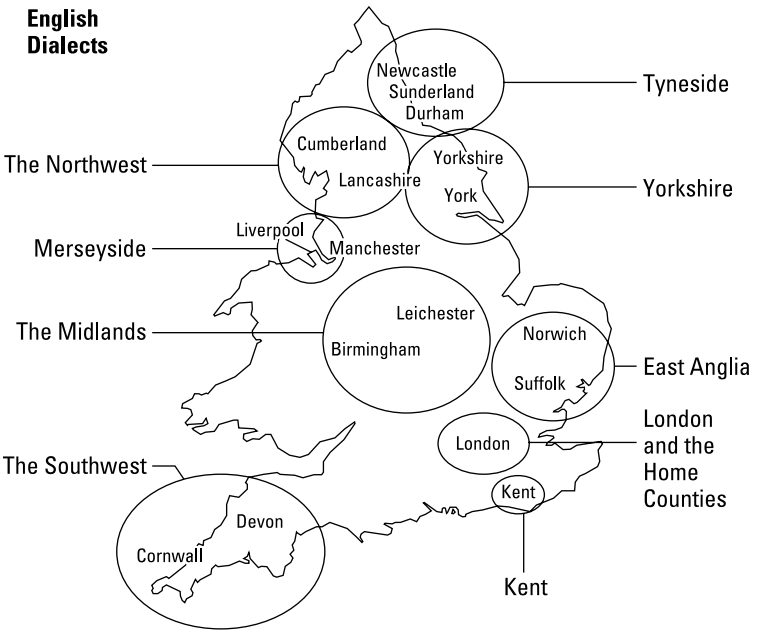
Estuary English refers to a new accent (or set of accents) forming among people living around the River Thames in London. However, before exploring this fine-grained English accent, let me start with some basics.

England is a small and foggy country, crammed with amazing accents. At the most basic level, you can define broad regions based on some sound properties. Here are three properties that some dialectologists begin with:

- **Rhoticity:** This characteristic focuses on whether an "r" is present or not after a vowel, such as in "car" and "card." Large areas of the north aren't rhotic, while parts of the south and southwest keep r-colored vowels.
- **The shift from /ʌ/ to /ʊ/:** In the south, /ʌ/ remains the same, while in the north it shifts to /ʊ/, such that "putt" and "put" are pronounced [pʰʌt] and [pʰʊt] in the south but [ʊ] in the north.
- **The shift from /æ/ to /ɑ/:** This division has an identical boundary to the preceding shift. For example, consider the word "bath" [bɑθ].

Dialectologists further identify regional dialect groupings within England. Although experts may differ on these exact boundaries and groupings, a frequently cited list includes the following (Figure 18-4 maps these regions):

- London and the Home Counties, including *Cockney* (check out the next section for more information on Cockney)
- Kent
- The Southwest (Devon and Cornwall)
- The Midlands (Leicester and Birmingham) or *Brummie*
- East Anglia (Norwich and Suffolk)
- Merseyside (Liverpool and Manchester) or *Scouse*
- Yorkshire
- The Northwest (Cumberland and Lancashire)
- Tyneside (Newcastle, Sunderland, and Durham), or *Geordie*



Map by Wiley, Composition Services Graphics

Figure 18-4: A map of England showing accent regions.

In addition to these large geographical regions, consider that most of the English population lives in cities. English cities show much greater accent variation than the countryside, largely due to sociolinguistic factors. Because approximately 15 percent of England lives in London (and many features of

London English have spread to other cities), London is a great place to study urban English accents.

Forming Estuary English from Cockney

Some dialectologists suggest that the gap between prestige and working class forms in England create the perfect scenario for rising classes to find something in-between, *Estuary English.* According to this view, Estuary English is a bold, new dialect in formation. However, noted British phonetician John C. Wells instead maintains the evidence shows various sound changes coming from working-class London speech, each independently spreading. That is, many types of mid-level social accents are forming in London. They're considered various types of *London accents.*

At any rate, to form Estuary English, follow these steps:

1. **Swipe these ingredients from Cockney:**

   See the section, "Talking Cockney" for more specifics about them.

   - /θ/-fronting
   - Glottal-stop insertion
   - /l/-vocalization
   - /h/-dropping

2. **Add an intrusive "r."**

   So "law and order" becomes [ˈlɔɹəŋ ɔɹdə].

3. **Mix in these expressions: "Cheers! There you go!"**

   Awroyt! You're in the Estuary.

Talking Cockney

Cockney is one of the more notable London accents and perhaps the most famous, representing London's East End. Cockney is an urban, social dialect at one end of the sociolinguistic continuum, with Received Pronunciation (RP) at the other. Nobody knows exactly where the word "Cockney" comes from, but it has long meant *city person* (as in the 1785 tale of a city person being so daft he thinks a rooster neighs like a horse).

Cockney has many lexical characteristics, including rhyming slang (*trouble and strife*, for wife) and syntactic features (such as double negation). At the phonetic level, Cockney is most known for the following characteristics with its consonants:

- θ**-fronting:** Pronouncing words that in Standard English are normally /θ/ as [f], such as "think" as [fɪŋk] or "maths" as [mɛfs].

- **Glottal-stop insertion:** Inserting a glottal stop for a /t/ in a word like "but" [bʌʔ] or "butter" [ˈbʌʔə].
- /l/**-vocalization:** Pronouncing the /l/ in a word like "milk" as [u] to be [ˈmiuk].
- **/h/ dropping:** Dropping the /h/ word initially. Pronouncing "head" as [ɛd] or [ʔɛd].

Note: Many of these features have now spread to most British accents.

Meanwhile, Cockney also exhibits the following characteristics with vowels:

- **/iː/ shifts to [əi]:** "Beet" becomes [bəiʔ].
- /eɪ/ **shifts to [**æɪ~aɪ**]:** "Bait" becomes [bæɪʔ].
- **/**aɪ**/ shifts to [**ɑɪ**]:** "Bite" becomes [bɑɪʔ].
- **/**ɔɪ**/ shifts to [**~oɪ**]:** "Choice" becomes [tʃʰoɪs].
- **/**uː**/ shifts to [**ə⁂**] or [**⁂ː**] a high, central, rounded vowel:** "Boot" becomes [bə⁂] or [b⁂ːʔ] where [⁂] is a rounded, central vowel.
- /aʊ/ **may be [**æə**]:** "Town" becomes [tˢæə̃n].
- **/**æ**/ may be [**ɛ**] or [**ɛɪ**]:** The latter occurs more before /d/, so "back" becomes [bɛk] and "bad" becomes [bɛːɪd].
- **/**ɛ**/ may be [**eə**], [**eɪ**], or [**ɛɪ**] before certain voiced consonants, particularly before /**d**/:** "Bed" becomes [beɪd].

Cockney has already moved from its original neighborhoods out toward the suburbs, being replaced in the East End by a more Multiethnic London English (MLE). This accent includes a mix of Jamaican Creole and Indian/Pakistani English, sometimes called *Jafaican* (as in "fake Jamaican"). A prominent speaker of MLE is the fictional movie and TV character Ali G.

Wales: Wenglish for fun and profit

Wales is a surprising little country. It harkens back to the post-Roman period (about 410 AD). Until the beginning of the 18th century, the population spoke Cymraeg (Welsh), a Celtic language (pronounced /kəmˈrɑːɪg/). The fact that Welsh English today is actually a younger variety than the English spoken in the United States is quite amazing.

Currently, only a small part of the population speak Welsh (about 500,000), although this number is growing among young people due to revised educational policies in the schools. Welsh language characteristics and the accent features of the local English accents have a strong interplay, resulting in a

mix of different Welsh English accents (called *Wenglish*, by some accounts).

Bend It Like Beckham (phonemes, that is)

British dialects are associated with sociolinguistic factors that arguably have more consequences than in North America. Because dialect has long marked social class, changing one's accent remains important to social climbing. On one end of the spectrum is Received Pronunciation (RP) (think royals, upper class, professionals) and Cockney on the other (London working class). In England it would be difficult to imagine candidates such as Jimmy Carter, Bill Clinton, or George W. Bush — who have strong regional, Southern states American accents — achieving any success for speaking reasons alone (although the British Prime Minister Gordon Brown spoke an English that frequently betrayed his Scottish heritage).

Many media personalities in England seem to be aiming for a middle ground between RP and Cockney. They exhibit the right amount of urban cockiness, but not too much. Perhaps to hit success in the British media world, it helps to bend your phonemes like English soccer star David Beckham. Check out the following for help: `www.youtube.com/watch?v=I2X9L5llhTQ`.

Or stir-fry them, like Jamie Oliver at `www.youtube.com/watch?v=jIwrV5e6fMY`.

Characteristics of Wenglish consonants include the following:

- **Use of the voiceless uvular fricative /χ/:** "Loch" becomes [ˈlɒχ] and "Bach" becomes [ˈbɒχ].
- **Dropping of /h/ in some varieties:** Wenglish realizes produces "house" as [aʊs].
- **Distinction between /w/ and /ʍ/:** "Wine" and "whine" become [waĩn] and [ʍaĩn].
- **Distinction between /yː/ and /ɪʊ/:** In "muse" and "mews" and "dew" and "due."
- **Use of the Welsh /ɬ/ sound, a voiceless lateral fricative:** "Llwyd" is [ɬʊɪd] and "llaw" is [ɬau].
- **Tapping of "r":** "Bard" is pronounced as [bɑɾd].

Characteristics of Wenglish vowels include the following:

- **Distinction of [iː] and [ɪə]:** As in "meet" ([miːt]) and "meat" ([mɪət]), and "see" ([siː]) and "sea"([sɪə]).
- **Distinction of [e], [æɪ], and [eɪ]:** As in "vane" ([vẽn])," vain ([væĩn]), and "vein" (veĩn).
- **Distinction of [oː] and [oʊ]:** As in "toe" ([toː]) and "tow" ([toʊ]), and "sole" ([soː]) and "soul"([soʊl]).

- **Distinction of [oː] and [oə]:** As in "rode" ([roːd]) and "road" ([roəd]), and "cole" ([k^hoːl]) and "coal" ([k^hoəl]).

One characteristic for suprasegmentals includes distinctive pitch differences, producing a rhythmic, lilting effect. This accent occurs because when syllables are strongly stressed in Welsh English, speakers may shorten the vowel (and lower the pitch) of the stressed syllable. For instance, in the phrase "There was often discord in the office," pitch may often fall from "often" to the "dis" of "discord," but will then rise again from "dis" to "cord." Also, the "dis" will be short, and the "cord" will be long. This pattern is very different than what's found in Standard English (British) accents.

Scotland: From Aberdeen to Yell

Scottish English is an umbrella term for the varieties of English found in Scotland, ranging between Standard Scottish English (SSE) at one end of a continuum to broad Scots (a Germanic language and ancient relative of English) on the other. Scots is distinct from Scottish Gaelic, a Celtic language closer to Welsh. Thus, Scottish people are effectively exposed to three languages: English, Scots, and Scottish Gaelic.

This rich linguistic mix leads to *code shifting*, when talkers move back and forth between languages, preserving the phonology and syntax of each. Social factors where Scotsmen (and women) tend to speak English more in formal situations or with individuals of higher social status also affect this shifting. This type of language shifting is called *style shifting*.

Key characteristics of Scottish English consonants include the following:

- **Varieties of "r" for alveolars:** Examples include the alveolar *tap* (rapid striking of the tongue against the roof of the mouth to stop airflow), such as "pearl" pronounced [ˈpɛr̥] and the alveolar trill (/r/), such as "curd" pronounced [kʌrd].
- **Velarized /l/:** An example includes "clan" pronounced [kɬæ̃n].
- **Nonaspirated /p/, /t/, and /k/:** For instance, "clan," "plan," and "tan" would be [kɬæ̃n], [pɬæ̃n], and [tæ̃n]. In contrast, the GAE pronunciation of these words would begin with an aspirated stop (such as [t^hæ̃n]).
- **Preserved distinction between the /w/ and /ʍ/:** An example would be the famous "which/witch" pair, [ʍɪʧ] and [wɪʧ].

- **Frequent use of velar voiceless fricative /x/:** An example includes "loch" (lake) pronounced as [ɬɔx], and Greek words such as "technical" as [ˈtɛxnəkəɬ].

Characteristics of Scottish vowels are

- **No opposition of /ʊ/ versus /uː/:** Instead, /ʊ/ and /u/ are produced as a rounded central vowel. Thus, "pull" and "pool" are both [pʉɬ].
- **The vowels /ɒ/ and /ɔ/ merge to /ɔ/:** For example, "cot" and "caught" are both pronounced /kɔt/.
- **Unstressed vowels often realized as [ɪ]:** For example, "pilot" is pronounced as [ˈpʌiɬɪt].

Ireland: Hibernia or bust!

The English language has a venerable history in Ireland, beginning with the Norman invasion in the 12th century and gathering steam with the 16th Century Tudor conquest. By the mid 19th century, English was the majority language with Irish being in second place.

Because of the stereotype of an Irish dialect, don't fall into the trap of thinking that all Irish English accents sound alike. Irish English has at least least three major dialect regions:

- **East Coast:** It includes Dublin, the area of original settlement by 12th century Anglo-Normans.
- **Southwest and West:** These areas have the larger Irish-speaking populations.
- **Northern:** This region includes Derry and Belfast; this region is most influenced by Ulster Scots.

Within these broad regions, the discerning ear can pick out many fine distinctions. For instance, Professor Raymond Hickey, an expert on Irish accents, describes *DARTspeak*, a distinctive way of talking by people who live within the Dublin Area Rapid Transit District.

Like anywhere, accent rivalry occurs. A friend of mine, Tom, from a village about 60 kilometers east of Dublin, was once ranting about the *Dubs* and *Jackeens* (both rather derisive terms for people from Dublin) because of their disturbing accent. Of course, when Tom goes to Dublin, he is sometimes called a *culchie* (rural person or hick) because of *his* accent.

Despite these caveats about the variable nature of Irish English accents, Hiberno-English does have some common characteristics for consonants:

- **Rhotic:** Some local exceptions exist.
- **Nonvelarized /l/:** For instance, "milk" is [mɪlk]. A recent notable exception is in South Dublin varieties (such as *DARTspeak*).
- **Dental stops replace dental fricatives:** For instance, "thin" is pronounced as [t̪ɪn], and "they" as [d̪e:].
- **Strong aspiration of initial stops:** As in "pin" [p^{h}ĩn] and "tin" [t^{h}ĩn].
- **Preserved distinction between the /w/ versus /ʍ/, similar to Scottish English:** For example, "when" as [ʍɛ̃n] and "west" as [wɛst].

Hiberno-English has the common characteristics for vowels:

- **Offglided vowels /eɪ/ and /oʊ/:** "Face" and "goat" have steady state vowels outside Dublin, so they're pronounced [fe:s] and [go:t].
- **No distinction between /ʌ/ and /ʊ/:** In "putt" and "put," both are pronounced as [ʌ].
- **Distinction between /ɒ:/ and /o:/ maintained:** In "horse" and "hoarse," they're pronounced as [hɒ:rs] and [ho:rs], though not usually in Dublin or Belfast.

Here are some common characteristics for suprasegmentals:

- **Gained syllable:** Some words gain a syllable in Irish English, like "film," pronounced [ˈfɪlə̃m].
- **Lilting intonation:** Irish brogue typifies much of the Republic of Ireland (Southern regions), different from the north where there is more falling than rising intonation.

Transcribing Other Varieties

English is the main language in the United Kingdom, the United States, Australia, New Zealand, Ireland, Anglophone Canada and South Africa, and some of the Caribbean territories. In other countries, English isn't the native language but serves as a common tongue between ethnic and language groups. In these countries, many societal functions (such as law courts and higher education) are conducted mainly in English. Examples include India, Nigeria, Bangladesh, Pakistan, Malaysia, Tanzania, Kenya, non-Anglophone South Africa, and the Philippines. In this section, I show you some tips for hearing and transcribing some of these accents.

Australia: We aren't British

Australian English has terms for things *not* present in England. For instance, there is no particular reason that anyone should expect the land of Shakespeare to have words ready to go for creatures like *wallabies* or *bandicoots*. What's surprising is how Australian English accents have come to differ from those of the mother ship.

The original English-speaking colonists of Australia spoke a form of English from dialects all over Britain, including Ireland and South East England. This first intermingling produced a distinctive blend known as *General Australian English*. The majority of Australians speak General Australian, the accent closest to that of the original settlers. Regionally based accents are fewer in Australia than in other world English accents, although a few do exist. You can find a map showing these stragglers (with sound samples) at `http://clas.mq.edu.au/voices/regional-accents`.

As the popularity of the RP accent began to sweep England (from the 1890s to 1950s), Australian accents became modified, adding two new forms:

- **Cultivated:** Also referred to as *received,* this form is based on the teaching of British vowels and diphthongs, driven by social-aspirational classes. An example is former Prime Minister Malcolm Fraser.
- **Broad:** This accent is formed in counter-response to cultivated, away from the British-isms, emphasizing nasality, flat intonation, and syllables blending into each other. Think Steve Irwin, Crocodile Hunter.

Here are some things you should know about Australian accents:

- Like many British accents, Australian English (AusE) is no*n-rhotic*, meaning "r" sounds aren't pronounced in many words (such as "c*ard"* and "lead*er"*).
- However, Australians use *linking-r* and *intrusive-r,* situations where "r" appears between two sounds where it normally wouldn't be produced. For example, an Australian would normally pronounce "tuner" without an "r" sound at the end ([ˈtjʉːnə]), but if a word beginning with a vowel follows that word, then the "r" does appear ([ˈtjʉːnəɹ æ̃mp]). This is an example of linking r. See Chapter 7 for more information on linking- and intrusive-r.
- The "r" is produced by making an /ɹ/ and a /w/ at the same time, with lips somewhat pursed.
- Phoneticians divide the AusE vowels into two general categories by length:

- *Long vowels* consist of diphthongs (such as /æɪ/) and tense monophthongs (such as the vowels /oː/ and /eː/).
- Short vowels consist of the lax monophthongs (such as /ɪ/). See Chapter 7 for more information on English tense and lax vowels.

Here are a couple of AusE vowel features to remember:

- **Realization of /e/ as [æɪ]:** "Made" sounds like [mæɪd]. This feature is so well known that it's considered a *Shibboleth*, a language attribute that can be used to identify speakers as belonging to that group.
- **Realization of /u/ as a high, central, rounded vowel, [ʉː]:** "Boot" sounds like [bʉːt].
- **Realization of /ɑ/ as [ɔ]:** "Hot" sounds like [hɔt].
- **Realization of /ɛ/ as [eː]:** "Bed" sounds like [beːd].

New Zealand: Kiwis aren't Australian

New Zealand accents are attracting much study because they're like a laboratory experiment in accent formation. New Zealand didn't have its own pronunciation until as late as the 19th century when some of the pioneer mining-town and military base schools began forming the first, identifiable New Zealand forms. Although the English colonial magistrates weren't exactly thrilled with these Kiwi creations, the accents held ground and spread as a general New Zealand foundation accent. Much like the three-way regional dialect split in Australia, cultivated and broad accents were later established as the result of RP-type education norms introduced from England.

New Zealanders also show influences from Maori (Polynesian) words and phrases, including *kia hana* (be strong), an iconic phrase used following the 2010 Canterbury earthquake.

In recent years, New Zealanders have undergone a linguistic renaissance, taking pride in their accents, noting regional differences (such as between the north and south islands), and often taking pains to distinguish themselves linguistically from other former colonies, such as Australia, South Africa, and the United States.

Some attributes of the Kiwi accent for consonants include the following:

- **Mostly non-rhotic, with linking and intrusive r, except for the Southland and parts of Otago:** For example, "canner" would be [ˈkɛ̃nə] (non-rhotic). Yet, a linking "r" would be found in "Anna and Michael,"

sounding like "Ann*er* and Michael" (see Chapter 7 for more information on linking and intrusive "r").

- **Velarized (dark) "l" in all positions:** For example, "slap" would be [sɬɛp].
- **The merger of /w/ and /ʍ/ in younger speakers, although still preserved in the older generation:** Thus, younger New Zealanders would likely pronounce both "which" and "witch" with [w], while their parents would use /ʍ/ and/w/ instead.
- **Possibly tapped /w/ and intervocalic /t/:** (*Intervocalic* means between two vowels; refer to Chapter 2.) For example, "letter" is pronounced ['ɬɛrə].

ome key characteristics for Kiwi vowels include the following:

- **Use of a vowel closer to /ə/:** A big difference with Kiwi English is the vowel in the word "kit." Americans use /ɪ/ (and Australians would use /i/), Kiwis use a vowel closer to /ə/. Thus, "fish" sounds like [fəʃ].
- **Move of /ɛ/ toward [e]:** "Yes" sounds like [jes].
- **Move of /e/ toward [ɪ]:** "Great" sounds like [gɹɪt].
- **Rise of /æ/ toward [ɛ]:** "Happy" sounds like ['hɛpɪ].
- **Lowering of /ɔː/ to [oː]:** The words, "thought," "yawn," and "goat" are produced with the same vowel, [oː]. Americans can have a real problem with this change. Just ask the bewildered passenger who mistakenly flew to Auckland, New Zealand instead of Oakland, California (after misunderstanding Air New Zealand flight attendants at Los Angeles International Airport in 1985).

South Africa: Vowels on safari

South African English (SAE) refers to the English of South Africans. English is a highly influential language in South Africa, being one of 11 official languages, including Afrikaans, Ndebele, Sepedi, Xhosa, Venda, Tswana, Southern Sotho, Zulu, Swazi, and Tsonga. South African English has some ocial and regional variation. Like Australia and New Zealand, South African as three classes of accents:

- **General:** Middle class grouping of most speakers
- **Cultivated:** Closely approximating RP and associated with an upper class
- **Broad:** Associated with the working class, and closely approximating the second-language Afrikaans-English variety

All varieties of South African English are non-rhotic. These accents lose post-vocalic "r," except (for some speakers) liaison between two words, when the /r/ is underlying in the first, so for example, "for a while" as [fɔɹə'ʍɑːɬ]. Here are some key characteristics of South African English consonants:

- **Varieties of "r" consonants:** They're usually post-alveolar or retroflex [ɹ]. Broad varieties have [ɾ] or sometimes even trilled [r]. For example, "red robot" [ɹɛ̟d ˈɹɐʊbət], where "robot" means traffic light.
- **No instrusive "r":** "Law and order" is [ˈloːnoːdə], [ˈloːwənoːdə], or [ˈloːʔə̃noːdə]. The latter is typical of Broad SAE.
- **Retained distinction between /w/ and /ʍ/ (especially for older people):** As in "which" ([ʍɪʧ]) and "wet" ([wet]).
- **Velarized fricative phoneme /x/ for some borrowings from Afrikaans:** "Insect" is [xoxə].
- **/θ/-fronting:** /θ/ may be realized as [f]. "With" is [wɪf].
- **Strengthened /j/ to [ɣ] before a high front vowel:** "Yield" is [ɣɪːɬd].
- **Strong tendency to initially voice /h/:** Especially before stressed syllables, yielding the voiced glottal fricative [ɦ]. For instance, "ahead" is [əˈɦɛ̟d].

Some attributes for vowels in South African English are

- **Monophthongized /aʊ/ and /aɪ/ to [ɑː] and [aː]:** Thus, "quite loud" is [kʰwaːt lɑːd].
- **Front /æ/ raised:** In Cultivated and General, front /æ/ is slightly raised to [æ̝] (as in "trap" [tʰɹæ̝p]). In Broad varieties, front /æ/ is often raised to [ɛ]. "Africa" sounds like [ˈɛfɹɪkə].
- **Front /iː/ remained [iː] in all varieties:** "Fleece" is [fliːs]. This distinguishes SAE from Australian English and New Zealand English (where it can be the diphthongs [ɪi~əi~ɐi]).

West Indies: No weak vowels need apply

Caribbean English refers to varieties spoken mostly along the Caribbean coast of Central America and Guyana. However, this term is ambiguous because it refers both to the English dialects spoken in these regions and the many English-based creoles found there. Most of these countries have historically had some version of British English as the official language used in the courts and in the schools. However, American English influences are playing an increasingly larger role.

As a result, people in the Caribbean code switch between (British) Standard English, Creole, and local forms of English. This typically results in some distinctive features of Creole syntax being mixed with English forms.

At the phonetic level, Caribbean English has a variety of features that can differ across locations. Here are some features common to Jamaican English consonants:

- **Variable rhoticity:** Jamaican Creole tends to be rhotic and the emerging local standard tends to be non-rhotic, but there are a lot of exceptions.
- **/θ/-interdental stopping:** Words like "*th*ink" are pronounced using /t/ and words like "*th*is" are pronounced using /d/.
- **Initial /h/ deleted:** "*H*omes" is [õmz].
- **Reduction of consonant cluster:** Final consonant dropped, so "miss*ed*" is [mis].

Some attributes for vowels are as follows:

- **Words pronounced in GAE with /eɪ/ (such as "face") are either produced as a monophthong ([**eː**]), or with on-glides ([**ie**]):** Thus, "face" is pronounced as [feːs] or [fies].
- **Words pronounced in GAE with /oʊ/ (such as "goat") are either produced as ([**oː**]), or with on-glides ([**uo**]):** Thus, "goat" is pronounced as [goːt] or [guot].

  This difference between monophthong versus falling diphthong) is a social marker — the falling diphthong must be avoided in English to avoid social stigma (if prestige is what the speaker wishes to project).
- **Unreduced vowel in weak syllables:** Speakers use comparatively strong vowels in words such as "about" or "bacon" and in grammatical function words, such as "in," "to," "the," and "over." This subtle feature adds to the characteristic rhythm or lilt of Caribbean English (for instance, Caribbean Creoles and Englishes are syllable-timed).

Chapter 19

Working with Broken Speech

In This Chapter

- Getting a deeper understanding of adult speech disorders
- Delving into the dysarthrias
- Working with common child language disorders
- Applying special IPA symbols, when needed

Sometimes adults and children have speech, hearing, or language disorders that prevent them from communicating. Health professionals who deal with these disorders focus on researching, diagnosing, and treating those individuals. (In Canada and North America, the study of speech, hearing, and language disorders is known as *speech language pathology and audiology* whereas in the other parts of the world, this field is known as *logopediatrics and phoniatrics* or *clinical phonetics*.) Because speech problems may be a telling first symptom of progressive neurological disease (such as ALS or Parkinson's Disease), other medical professionals also need to understand these disorders.

At a basic human level, such problems should be of interest to anyone who has a family member with such ailments. For example, people who have family members in stroke clinics often complain that their loved ones don't get the kind of care they need because no one can understand their loved one's speech. Tuning in to disordered speech by means of spectrographic evidence (as I discuss in Chapter 13) and narrow transcription, as this chapter explains, are good ways to better understand the nature of these individuals' speech difficulties.

Transcribing Aphasia

Aphasia is a language disorder in adults resulting from brain injury or disease. Depending on where the damage is located in the brain and how extensive it is, the person may experience very different symptoms. Most classification systems agree on a series of aphasic syndromes, based on a

profile of speaking and listening abilities. The two most common syndromes are Broca's aphasia and Wernicke's aphasia, named after two famous 19th century scientists.

Transcribing the speech of these different aphasic syndromes presents very different challenges because of the quantity and quality of speech you will work with. These sections show you sample transcriptions of individuals with these disorders.

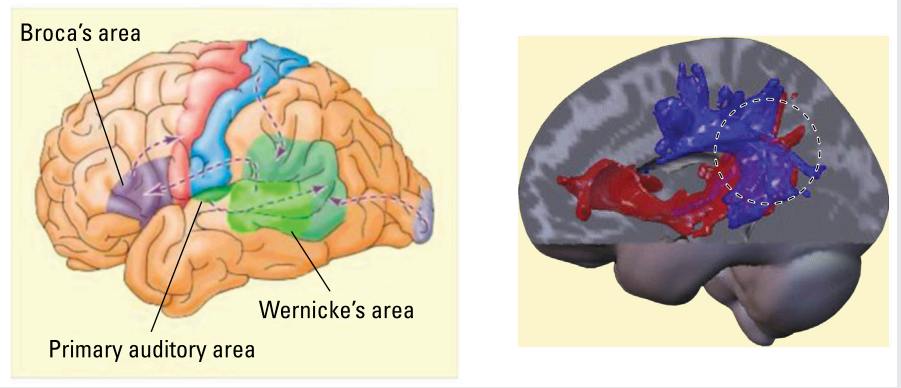
Broca and Wernicke: Lasting insights in aphasiology

The study of aphasia owes a tremendous debt to two geniuses from Europe whose insights have stood the test of time. Pierre Paul Broca (1824–1880) was a French neurosurgeon who studied individuals with brain damage and compared their neuroanatomy to their speech and language output. Because he (obviously) had no CT or MRI scanners available at the time, Broca examined a dead patient's brain *(post-mortem*), looked for lesions (he called *softenings*), and hypothesized how this damage might affect speech and language. From this painstaking work, he identified a part of the left frontal lobe (which is now known as *Broca's area (BA),* that he considered responsible for articulated language.

This brain discovery was an important anatomical proof of *localization of function*, the idea that a part of the brain could be responsible for a particular type of behavior. Broca's work also associated speech production behavior to the left side of the brain — thus providing key evidence of brain *lateralization*, that the brain will use one side differently than the other.

Early understanding of the neural basis of speech became more sophisticated with the contributions of Karl Wernicke (1848-1905). Wernicke used similar research techniques as Broca, but came up with a rather different view. This Prussian-born German neurologist is most famous for another part of the brain, called *Wernicke's area, (WA),* located at the top, back part of the brain's left temporal lobe. People with damage in this area could speak fluently, but their speech was often jumbled and didn't make sense.

Putting everything together, Wernicke proposed a model in which there is a sort of loop (or connection) between WA (where speech sounds are heard and decoded) and BA (associated with the production of speech and language). The dynamic nature of this model (with information flowing from one site to another) gave rise to interesting predictions: For example, there should be fiber connections between WA and BA, and its damage should cause particular problems with repetition. Years later, such a structure, *the arcuate fasciculus* (meaning *arc-like bundle*) was identified. Evidence suggested problems with repetition occurring when it was damaged (although it now seems that neural tissue feeding the *arcuate fasciculus*, rather than the fasciculus itself, can account for the repetition problems, leading to proposed refinements of Wernicke's model). Refer to this figure to see the brain showing BA, WA, and the relation of these areas to the primary auditory area (where basic speech sounds are processed).



The insights of these fathers of aphasiology have been refined and added to, but have generally stood the test of time. One recent addition is located in parietal (side) cortex, called *Geschwind's territory* (see the circled area in the right figure), which seems to have a role in regulating speech as people hear themselves talk (motor control during perception). For more information on the brain bases of speech and language, see "The Brain from Top to Bottom" a site developed by Bruno Dubuq funded by the Canadian Institutes of Health Research at `http://thebrain.mcgill.ca/index.php`.

Broca's: Dysfluent speech output

Broca's aphasia is most commonly caused by damage to the left, frontal part of the brain. It results in halting, choppy speech that has poor *melody* (speech frequency and rhythm qualities). Depending on severity, the patient may be able to produce words and phrases, or almost nothing at all (sometimes called being at the *one word stage*). Patients have particular difficulty with words that are part of the grammar, called *closed class* or *function words,* which includes word endings that carry meaning (such as "-ed" or "-s"), common determiners and prepositions ("a," "the," "to," "over," and so on), and pronouns ("he," "she," "it," "they," and so on). They may leave out or poorly produce difficult words.

The following is a short transcribed speech sample from an individual with Broca's aphasia.

> "I'm no good. Um. Ache(s). And . . . a. a. a. home. (A) doctor. And legs. Walking no good."
>
> [|æm˥no ˈgʊd ə|ə̃m| eɪk(s)|æn˥d‖ə/ə/ə|əˈhõm|əˈdɑktəˈ|æn˥d lɛgz|ˈwɑkĩŋ no gʊd‖]

Wernicke's: Fluent speech output

The Wernicke's aphasic patient presents different challenges for transcription than the Broca's aphasic speaker. Rate, intonation, and stress are usually normal. Because speech is often plentiful, getting a sufficient *corpus* (body of speech to analyze) likely won't be a problem, as is often the case for *dysfluent* (halting, disrupted) speech. However, trying to understand words can be difficult at times because you, the listener, may simply have no idea what your subject is talking about.

In more extreme cases, patients may show *press for speech* (talking rapidly and interrupting others), or *logorrhea* (rambling, incoherent talkativeness). If you're gathering a corpus under such circumstances, experienced clinicians recommend using gentle but firm affirmations such as "Yes, I know" or "You are right. I got it" to wrest back control of the interviewing situation.

In Wernicke's aphasia, word errors are commonly *paraphasic,* when unintended syllables, words, or phrases intrude during the effort to speak. Fluent aphasics have many more paraphasic errors than nonfluent (Broca's type) aphasics. These paraphasic errors can involve the substitution of one word for another, called *verbal paraphasias* (like "bug" for the target "bun"). When a production is unrecognizable because more than half is produced incorrectly, it's called a *neologism* (made up word), such as "weather" realized as "belimmer."

Here is an example transcription of the speech of an individual with Wernicke's aphasia.

> "Oh, about uh . . . about a hundred and . . . let's see, a hundred and . . . thirty. About forty."
>
> [|o baʊʔ tʰə| baʊʔ ə ˈhɐ̃ndˠɹɛd ɛ̃n|lɛts sɪ ə ˈhɐ̃ndˠɹɛd ɛ̃n|ˈðɪɹɪ| əˌbaʊʔ ˈfɔɹi‖]

Dealing with phonemic misperception

A challenge in working with the speech of people with speech disorders, such as Broca's aphasia and apraxia of speech (AOS) (which I discuss later in this chapter) is phonemic misperception. *Phonemic misperception* happens when your subject intends to produce a certain speech target but instead makes an error from improper timing or coordination. As a result, you (the listener) don't know into which perceptual sound category the production should fall.

Remember, you're hearing many of these sounds categorically. Did he mean "*s*ee" or "*sh*e"? Did he mean to say "*p*en" or "*Be*n"?

One of the reasons that family members of patients with Broca's aphasia and/or AOS report understanding them better than other people could be that they're relying on other information (such as body language or other contextual cues). However, knowing the root of these patients' problems can help you better understand the situation. Here are two points to remember:

- **Damage to the posterior parts of the brain's speech area, such as in Wernicke's aphasia, results in sound selection errors.** A *sound selection error* is when an intended sound is misselected, resulting in the wrong sound being chosen.

  So if a patient with Wernicke's aphasia makes an error saying the word "*p*en" (that you hear as "*B*en"), the chances are he has produced a well-formed /b/ because this speech error likely took place at a selectional level, higher up in the system. When it came time to map the object (a pen) into a word, he chose the wrong phoneme, accessing a well-produced, but wrong, sound.

- **Damage to the anterior parts of the brain's speech area, such as in Broca's aphasia or AOS, results in sound implementation errors.** In sound implementation errors, the intended sounds are correctly chosen higher up in the system (at a phonemic level). A breakdown occurs when the patient's brain sends this information to the speech articulators.

  This type of patient correctly choses the phonemes /p/, /ɛ/, and /n/ for speech output. However, after selection, the initial phoneme becomes mistimed and uncoordinated while speaking. As a result, its timing properties (such as voice onset time) no longer fit in the nice neat categories that you're waiting for. It ends up sounding like a "b" (although perhaps not as clear as the one produced by the Wernicke's aphasic).

Figure 19-1 shows a flowchart of selection and implementation errors. This figure shows two possible routes for producing an apparent sound substitution error by an aphasic talker. The patient sees a pen, activating the correct concept ("pen") and a concept starting with a similar phoneme, "Ben." In a sound selection error, as in Wernicke's aphasia (shown by the dotted line) the patient selects the wrong item at a phonemic level, /b/, then correctly outputs this sound. In an implementation error, as in Broca's aphasia (shown by the solid line) the correct phoneme is selected, /p/, however this choice is then distorted or mistimed such that the final output sounds like [b].

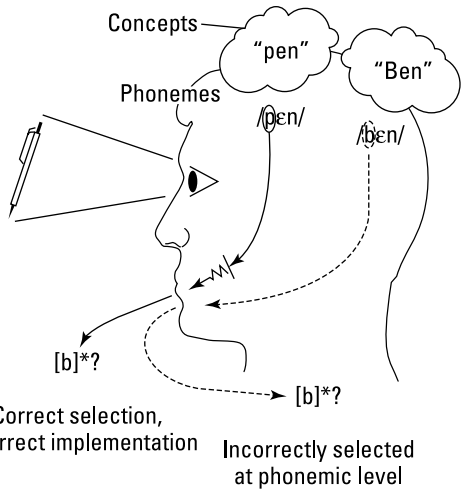


Figure 19-1: Selection versus implementation errors.

Using Special IPA to Describe Disordered Speech

Depending on the level of detail needed, you can find anything from broad (phonemic) transcription to more narrow description (including some allophonic variation) in clinical practice. An extension of the IPA has been developed to provide additional detail for disordered speech. A group of linguists interested in transcribing disordered speech started this system, called *ExtIPA*, in 1989. Since that time, phoneticians have also used the ExtIPA symbols to indicate sounds that come up during transcription of healthy speech, such as hushing, gnashing teeth, and smacking lips.

Figure 19-2 lists these special symbols that phoneticians who work with disordered speech use.

The top of Figure 19-2 in the area that I've labeled No. 1 shows features for consonants organized by manner (rows) and place (columns) of articulation. As in the regular IPA chart (refer to Chapter 3 for more information), voiced and voiceless sounds are listed side by side. A few things are different here than the regular IPA.

Take a few moments and take a close look at Figure 19-2. Unlike the regular IPA, the ExtIPA provides detail for abnormalities of airflow, phonation, nasalization, articulatory strength, articulation, and timing. The ExtIPA also lists a way to mark different prosodic modes (such as falsetto and loud speech) using musical expressions. For instance, look at the *nareal fricatives* area, which is making a hissing sound by blowing out the *nares* (nasal holes) of the nose. The percussive row also allows for lip smacks [ʬ] and teeth gnashes [ʭ].

1 CONSONANTS (other than on the IPA Chart)

	bilabial	labiodental	dentolabial	labioalv.	linguolabial	interdental	bidental	alveolar	velar	velophar.
Plosive		p̪ b̪	p̄ b̄	p̪͆ b̪͆	t̪ d̪	t̪̄ d̪̄				
Nasal			m̄	m̪͆	n̪	n̪̄				
Trill					r̪	r̪̄				
Fricative median			f̄ v̄	f̪͆ v̪͆	θ̪ ð̪	θ̪̄ ð̪̄	h h̄			fŋ
Fricative lateral+median								ɬ ɮ		
Fricative nareal	m̃̆							n̆̃	ŋ̆̃	
Percussive	ʬ						ʭ			
Approximant lateral					l̪	l̪̄				

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

3 DIACRITICS

̼	labial spreading	s̼	ʷ	strong articulation	sʷ	̃	denasal	m̃
̪	dentolabial	v̪	˕	weak articulation	v˕	̃	nasal escape	ṽ
̺	interdental/bidental	n̺	\	reiterated articulation	p\p\p	̴	velopharyngeal friction	s̴
̻	alveolar	t̻	↑	whistled articulation	s↑	↓	ingressive airflow	p↓
̼	linguolabial	d̼	→	sliding articulation	θ→s	↑	egressive airflow	!↑

2 CONNECTED SPEECH

(.)	short pause
(..)	medium pause
(…)	long pause
f	loud speech [{$_{f}$laud$_{f}$}]
ff	louder speech [{$_{ff}$laudə$_{ff}$}]
p	quiet speech [{$_{p}$kwaiət$_{p}$}]
pp	quieter speech [{$_{pp}$kwaiətə$_{pp}$}]
allegro	fast speech [{$_{allegro}$fast$_{allegro}$}]
lento	slow speech [{$_{lento}$slow$_{lento}$}]
crescendo, ralentando, etc. may also be used	

4 VOICING

ˬ	pre-voicing	ˬz	
ˬ	post-voicing	zˬ	
̥	partial devoicing	z̥	(ð̞)
̊	initial partial devoicing	z̊	(ð̞)
̥	final partial devoicing	z̥	(ð̞)
̬	partial voicing	s̬	(t̬)
̬	inital partial voicing	s̬	(t̬)
̬	final partial voicing	s̬	(t̬)
̄	unaspirated	p̄	
ʰ	pre-aspiration	ʰp	

OTHERS

(‾), (⏑)	indeterminate sound, consonant	(())	extraneous noise	((2 sylls))
(V̄), (P̄L̄V̄L̄S̄)	indeterminate vowel, voiceless plosive, etc.	!	sublaminal lower alveolar percussive click	
(N̄), (V̄)	indeterminate nasal, probably [v], etc.	ǃ	alveolar and sublaminal clicks (cluck-click)	
()	silent articulation (f), (m)	*	sound with no available symbol	

Figure 19-2: The ExtIPA symbols.

The section in Figure 19-2 marked No. 2 provides an astounding array of diacritics, to cover anything from whistled articulation, indicated with an up-arrow under a symbol [↑], to denasalization, such as you may have made while being stuffed up with a head cold. Denasalization is indicated by a tilde with a slash through it [̴].

A third section in Figure 19-2 labeled No. 3 deals with connected speech, including three lengths of pauses and four levels of volume. A fourth section labeled No. 4 provides an interesting array of choices to describe voicing. In addition to voiced, voiceless, and aspirated (states of the glottis that I cover in Chapter 2), the ExtIPA allows you many different partial states. The one most important here for clinicians is *unaspirated* (not having a puff of air after a stop consonant burst), indicated by an equal sign placed to the upper right of a phoneme, such as [pʰ]. Missing aspiration for syllable-initial voiceless stops is a common feature, requiring notation in clinical transcription. This equal sign diacritic for the feature unaspirated is actually an old diacritic that used to be in common clinical usage, which has apparently been revived.

Some of the ExtIPA symbols are occasionally used to transcribe everyday normal speech sounds in certain languages. For example, the diacritic linguoalabial (looking like a little seagull [̼]) turns out to be a regular feature of the Polynesian language Vanuatu. To make a linguolabial sound, place your tongue tip or blade against the upper lip and then release.

Referencing the VoQS: Voice Quality Symbols

The ExtIPA doesn't include symbols used for voice quality, such as whispering, creaky voice, or *electrolarynx* speech (made with a mechanical buzzing device, usually after vocal fold surgery). Therefore, a group of phoneticians devised a series of voice quality symbols (VoQS).

These symbols allow a phonetician to mark whether a healthy person starts whispering (indicated with two dots under the voiced symbol) or yawning (a raising symbol for *open jaw voice*). This list includes provisions to cover speech while the tongue is protruded (I am assuming pathology here) and substitute situations for a *pulmonic egressive airstream* (outflowing air from the lungs), including the use of *oesphogeal* and *tracheophageal* speech (a kind of burping speech that patients may be taught to permit speaking after *laryngectomy,* the surgical removal of the larynx and vocal folds). See to Figure 19-3 for the VoQS.

Airstream Types

Œ	oesophageal speech	ʢ	electrolarynx speech
ʡ	tracheo-oesophageal speech	↓	pulmonic ingressive speech

Phonation Types

V	modal voice	F	falsetto
W	whisper	C	creak
V̤	whispery voice (murmur)	V̰	creaky voice
Vʰ	breathy voice	V̝	whispery creak
V!	harsh voice	V!!	ventricular phonation
V̰!!	diplophonia	V̤!!	whispery ventricular phonation
V̪	anterior or pressed phonation	W̠	posterior whisper

Supralaryngeal Settings

L̝	raised larynx	L̞	lowered larynx
Vʷ	labialized voice (open round)	Vʷ	labialized voice (close round)
V̫	spread-lip voice	Vᵛ	labio-dentalized voice
V̺	lingo-apicalized voice	V̼	linguo-laminalized voice
V˔	retroflex voice	V̪	dentalized voice
V̡	alveoloarized voice	V̡ʲ	palatoalveolarized voice
Vʲ	palatalized voice	Vˠ	velarized voice
Vʁ	uvularized voice	Vˤ	pharyngealized voice
V̴ˤ	laryngo-pharyngealized voice	Vᴴ	faucalized voice
Ṽ	nasalized voice	Ṽ	denasalized voice
J̟	open jaw voice	J̠	open jaw voice
J̹	right offset jaw voice	J̺	left offset jaw voice
J̽	protruded jaw voice	ϴ	protruded tongue voice

Figure 19-3: The voice quality symbols (VoQS).

Transcribing Apraxia of Speech (AOS)

Apraxia refers to problems understanding or performing an action in response to a verbal command or in imitation. There are many types of

apraxias, including *buccofacial apraxia,* in which patients have difficulty moving the lips, tongue, and jaw when requested or shown.

The apraxias are interesting disorders. For instance, some patients in our clinic (at the University of Texas at Dallas) with buccofacial apraxia can't blow out a candle if asked. They may try something close (like opening their mouth or saying "blow"). However, if a clinician lights a match and holds it up near the patient's lips, the patient can usually blow it out just fine. In such a case, different neural regulatory systems are presumed to operate.

In apraxia of speech (AOS), also known as *verbal apraxia,* patients have effortful, dysfluent speech marked by many speech errors. (In other words, they struggle to get their speech out and make many mistakes.) Their word errors are typically *literal paraphasias,* where the patient produces more than half of the intended word. For example, a patient may say /ki/ instead of /ski/. Switching sounds, also called sound transposition, can also occur, such as "bukertup" for "buttercup."

Although there are documented cases of individuals with isolated AOS, this disorder is usually *comorbid* (occurs along with) with Broca's (nonfluent) aphasia. As a result, clinicians and researchers are challenged to isolate the higher-order language components from speech motor processing involved in these individual's errors.

Here you can see a short transcription of an American male speaker with mild-to-moderate AOS. This patient is describing the "Cookie Theft Picture," from the Boston Diagnostic Aphasia Exam, a well-known diagnostic test for aphasia.

> "Wo-man . . . uh . . . uh . . . washing. Uh. Bo-Uh baby, baby not. Boy. Mmmm . . . juh- uh jip- jip- [meaning: trip] no. Thister, sister. Uh party no p-party heh not. Pappy? No!"
>
> [|ˈwʊ.mə̃n| ə/ə|ˈwɑʃ̃ŋ̩| {$_{ff}$ bo $_{ff}$} ə ˈbebɪ|ˈbebɪ nɑʔt| bɔɪ| m̩| dʒə. ə |dʒɪʔp/ dʒɪʔp| no|ˈθɪstɚ ˈsistɚ|ə ˈpɑɹi no p/pɑɹɪ hɛ̃nɑʔ|ˈpʰæpi| no‖]

In this transcription, you can see some typical features of AOS while also getting an idea of how a transcription might handle these features. The patient shows a pause between the syllables of "wo" and "man" in the first word ("woman"). This syllable-timed, scanning speech pattern (typical of AOS) is indicated by using a dot between the syllables, marking a syllable division. Stuttered syllables (such as [ə] and [dʒɪʔp]) are indicated with slash marks, following the ExtIPA. As the patient tries to say "baby," a paraphasic production "bo" comes out loudly. This loudness is indicated with brackets and "ff" marks, following ExtIPA conventions. There are other substitution errors, such as "thister" for sister. From even this brief corpus, you can tell the patient knows he isn't expressing his intended meaning.

Transcribing Dysarthria

Dysarthria is the most frequently reported speech motor disorder. It refers to a group of speech disorders resulting from a disturbance in neuromotor control. It's typically speech distortion, rather than a problem of planning or programming. It results from problems with the speed, strength, steadiness, range, tone, or accuracy of speech movements. Dysarthria can affect articulation, phonation, respiration, nasality, and prosody. It can affect the clarity of speech and the effectiveness of spoken communication.

Dysarthria can affect children (such as in cerebral palsy and cases of childhood stroke or traumatic brain injury) as well as adults. In adults, common causes include traumatic brain injury, stroke, and progressive neurological diseases (Parkinson's disease, MS, ALS). This section provides some discussion of cerebral palsy, Parkinson's disease, and ataxic dysarthric speech.

Cerebral palsy

People with cerebral palsy have speech problems resulting from difficulties with muscle tone, reflexes, or motor development and coordination. Chapter 13 provides more information on this disorder, including a spectrogram.

Challenges in transcribing speech produced by individuals with cerebral palsy include problems associated with poor breath support, laryngeal and velopharyngeal dysfunction, and oral articulatory problems. Speech can suddenly be loud, resulting in distorted recording. Excess nasality can make judgments on certain consonants difficult. Starting and stopping at places other than usual phrase breaks can contribute to distorted *prosody* (language melody) and difficulty with word endings.

Here is a sample transcription from dysathric speech produced by a woman with CP. She is reading sentences from the Assessment of Intelligibility of Dysarthric Subjects (AIDS) test battery. In this corpus, you can observe false starts, difficulty with word endings, and many consonant and vowel distortions.

> "The canoe floated slowly down the river"
>
> [|dɛ̃ kˣ̄ɔ̃nu fo̰ʔɪʔ| {$_{f}$ foɪt| (ʔ) l̩. lʊɬɪ$_{f}$}|da̰ʊ̰a ˈɹɪvə‖]

The diacritic [ˣ̄] indicates lack of aspiration, and [˜] indicates nasal escape during a vowel. This subject also had a burst of loud speech, marked by the brackets {ff}.

Parkinson's disease

Parkinson's disease (PD) is a progressive movement disorder, meaning that symptoms continue and worsen over time. It results from the malfunction and death of important nerve cells in a part of the brain called the *substantia nigra* (black body), which secretes *dopamine,* a chemical that helps the brain control movement and coordination. As PD progresses, a person receives less and less dopamine and has increasing difficulty with movement control.

Individual symptoms vary rather widely from person to person. However, the primary motor signs of PD include

- Tremor of the hands, arms, legs, jaw, and face
- Rigidity or stiffness of the limbs and trunk
- Slowness of movement
- Impaired balance and coordination

Lionel Logue: A pioneer speech therapist and the quest to treat stuttering

The life of Australian speech therapist Lionel Logue was featured in the 2010 historical film, *The King's Speech*. The movie emphasized Logue's role as the personal speech consultant to King George VI of England. Logue helped the king overcome his stuttering by using a variety of ingenious and compassionate methods and by creating a close personal rapport with the king.

Two questions that people frequently ask include "Was Logue really like that?" and "Can people really cure stuttering like that?"

According to Caroline Bowen, an expert on Logue and advisor to the film producers, much of what was shown in *The King's Speech* is likely an accurate portrayal of Logue's practice. Logue was an elocutionist and specialized in "speech defects." However, no historical record of Logue's actual methods exist, and the methods shown in the movie were therefore subject to dramatic interpretation. He used his intuition and skills as a teacher and coach to help people with speech problems. He may not have had today's most current methods, but somehow his system seems to have worked. Logue was responsible for co-founding what eventually became the Royal College of Speech Language Therapists (RCSLT) in the U.K.

This drama did, however, have a bit of Hollywood in it. Logue probably didn't have his subjects curse and swear as in the movie, nor is there any record of Logue having made the king allow himself to be called "Bertie" by Logue. Nevertheless, it seems the movie was fairly close to the real Lionel Logue, a true pioneer speech therapist.

The question about a cure for stuttering is more difficult. At the time of Logue's practice, *stuttering* (also called *stammering*) was understood to be a psychological problem and the object of shame. There was no understanding of its biological basis.

Today, stuttering is defined as a speech disorder in which sounds, syllables, or words are repeated or last longer than normal. It goes under the broader category of speech *dysfluency* problems.

Symptoms include repeating consonants, words, parts of words, or phrases ("I got . . . I got my desk" "I . . . I know it" or "Mu-mu-mu-must"). Stuttering may include vocal spasms and a forced, almost explosive sound to speech. The person may appear to be struggling to speak. Hesitations can occur, including sound prolongations ("She is Dooonnna Jones") and *interjections* (putting in extra sounds or words) ("I got . . . uh . . . my book"). Body language that can accompany stuttering includes eye blinking, jerking of the head and other body parts, and jaw jerking. Here are some things you should know about stuttering:

- Stuttering tends to run in families. Genes that cause stuttering have been identified.
- About 5 percent of children aged 2-5 will develop some stuttering during childhood, which may last for several weeks to several years.
- Problems that persist or worsen in young children are called *developmental stuttering,* the most common type of stuttering.
- Stuttering can also result from brain injuries, such as stroke. This type is called *acquired neurogenic stuttering*.
- In rare cases, stuttering may be caused by emotional trauma (called *psychogenic stuttering*).
- Stuttering is more common in boys than girls. It also tends to persist into adulthood more often in boys than in girls.
- Stressful social situations and anxiety can make symptoms worse.
- Some people who stutter find that they don't stutter when they read aloud, sing, or whisper.

What would Lionel Logue do for a stutterer today? There is no magic cure for stuttering dysfluency, because stuttering is a complex problem that requires a comprehensive approach to treatment. You can find an up-to-date series of treatment guidelines maintained by the American Speech and Hearing Association (ASHA) at `www.asha.org/policy/GL1995-00048.htm`.

Most practitioners now use behavioral methods to help reduce the severity, duration, and abnormality of stuttering behaviors until they resemble normal speech, including a variety of techniques, such as modeling sounds and practice, working on slowed rate and control, incorporating relaxation exercises, and introducing repair strategies.

Amazingly, the neurological basis of stuttering remains a mystery. One hypothesis currently being explored is that stuttering may result from an over-reliance on feed-forward (as opposed to feedback) processing. Chapter 4 gives more information on feed-forward and feedback processes in speech. Stutterers receive tremendous benefit from *choral* (or *unison*) repetition, speaking at the same time as others. This strong effect has been documented for years, (see `www.youtube.com/watch?v=Xw_rVGUXgos` for a demo) and has recently led to the development of some instrumental approaches for treatment.

A new frontier being considered in stuttering research is drug treatment. For instance, olanzapine (Zyprexa) is an atypical anti-psychotic drug that blocks dopamine receptors. In a test of 24 adult stutterers conducted over 12 weeks, stuttered syllables decreased by 33 percent in the subjects taking the medication (and 14 percent for the subjects taking a placebo). These results suggest that drugs working on dopaminergic pathways may have a future in stuttering treatment, at least for some patients.

Scientists estimate that 89 percent of people with PD have speech and voice problems. Scientists think these problems result from inadequate merging of *kinesthetic feedback* (the feeling of the tongue, mouth, lips, and jaws) motor output and *context feedback* (hearing one's self talk). Other problems include abnormal sensory processing (feeling, tasking, seeing) and an impaired ability to initiate a motor response (getting a movement started).

The speech of people with PD is typically called *hypokinetic dysarthria* because scientists think that an undershooting of articulatory movements mark it. (In other words, for these patients the tongue, lips, and jaw don't move as much as they think they do.) Such speech is characterized by reduced loudness, monotonous pitch, reduced stress, imprecise articulation, short rushes of speech, breathy hoarseness, and hesitant and dysfluent speech.

Here is a sample transcription of an 84-year-old woman who has had PD for 22 years. Because she was *hypophonic* (low voice volume), the transcriber was unable to determine what was said in many instances, which is typical for speech of individuals with advanced PD.

> "But when I look at that, for in(stance?), that sign . . . when I look I get double vision that far. It's better on this side. Eyes are better, too."
>
> [{$_{pp}$|bə ɱɛ̃naɪ ˈlʊkɪʔ θæʔ fɔɹ̃ĩn|ðæ ˈsaĩn|wɛ̃naɪ lʊk aɪ gɛt ˈdʌbl̩ ˌvɪʃn̩ ðæ fɑɹ|ɪs bɛdɹ õn ðɪs saɪd|ˈaɪzə˞ ˌbɛɾɹ̩ t^{h}u ‖$_{pp}$}]

ExtIPA bracketing {$_{pp}$} notes that the speaker used low volume throughout.

Ataxic dysarthria

Ataxic (without ordered movement) dysarthria is an acquired neurological speech deficit thought to result from problems with the *cerebellum,* a part of the brain that regulates speech motor programming and fine motor execution. Abnormalities in articulation and prosody are hallmarks of this disorder. Typical problems include abnormalities in speech modulation, rate of speech, explosive or scanning speech, slurred speech, irregular stress patterns, and mispronounced vowels and consonants.

Here is a transcription of a 60-year-old male with *olivopontocerebrellar degeneration,* a disease that causes areas deep in the brain, just above the spinal cord to shrink. This progressive neurological disease affected his gait, motor control, and speech, leaving him with ataxic dysarthria.

> "And I do have one child that was a professor uh in college for a while and but right now she is working for Cisco."
>
> [|ãɛ̃n˥d|aɪ ˈdo hæv |wɑ̃n˥tʃɪɬd ðæʔ wʌz ə pɹəˈfɛ.sə˞|a̲ ĩn ˈk^{h}ɑ.lɪdʒ fɔɹ əˌɱaɪ!ɬ ãɛ̃n|bəʔ ɹaɪʔnaʊ ʃiəz wɜ˞kĩŋ fə˞ ˈs^{h}ɪ̤skɔ‖]

VoQS symbols are used here: harsh voice [!], creaky voice [̰], and breathy voice [̤]. Also, the [s] of "Cisco" is marked with an aspiration diacritic ([^{h}]) to show this consonant was made extra breathy.

Introducing Child Speech Disorders

Any parent who has had the thrill of hearing a child's first word can imagine the disappointment and worry that goes with the child having speech and language disorders. Because such disorders occur in a developing child, whose speech and language is growing along with other skills (including social and cognitive), coming up with a clear definition of such disabilities has been surprisingly complex and difficult.

A number of issues can contribute to speech and language problems in children. They can include:

- Hearing loss
- Language-based learning difficulties
- Neglect or abuse
- Intellectual disability
- Neurological problems, such as cerebral palsy, muscular palsy, muscular dystrophy, and traumatic brain injury, which can affect the muscles needed for speaking
- Autism
- Selective *mutism* (when a child won't talk at all in certain situations, often at school)
- Structural problems, such as cleft lip or cleft palate
- Childhood apraxia of speech (CAS), a specific speech disorder in which the child has difficulty in sequencing and executing speech movements
- Specific language impairment (SLI)

For more details, please consult `www.asha.org/public/speech/disorders/childsandl.htm`.

These next sections describe some of the basic speech problems that clinicians note in healthy children and compare these processes with the types of disorders noted in children with childhood apraxia of speech (CAS).

Noting functional speech disorders

In clinical practice, many speech language pathologists working with children classify a series of problems known as *functional misarticulations* also referred to as *functional speech disorders*. When a child suffers from one of these disorders, he or she has difficulty learning to make a specific speech sound (such as /ɹ/), or a few specific speech sounds, typically involving the following fricatives and approximants: /s/, /z/, /ɹ/, /l/, /θ/, and /ð/.

The difficulty with a group of predictable sounds is different than overall sound sequencing impairments (childhood apraxia of speech) or with slurring or problems with general motor control (dysarthria).

Some of these difficulties are commonly known, such as *lisps* (producing an intended /s/ as [θ]) and labialization of rhotics (intended /ɹ/ realized as [w]). For instance, clinicians commonly encounter errors such as "willy" or "thilly" (for "really" or "silly"). Clusters are reduced (such as "spill" being realized as "pill"). Syllable-final consonants may be deleted, such as "fruit" being realized as "fru". Substitution includes fronting (such as "king" becoming "ting") and stopping (such as "bath" becoming "bat").

If a clinician observes a less usual pattern of sound production errors in children (such as the intrusion of glottal stops, substitutive backing, sound switching errors, or initial consonant deletion), the clinician will probably suspect a different underlying disorder.

Here is a transcription of a 3-year-old girl with functional speech disorder. She is a monolingual speaker of English describing a sticker animal to a researcher.

> "Do you know what my teacher first did? Guess? But this one ripped!"
>
> [|dɛ jũ no wʌʔ maɪ ˈdɪʤəfɚʔ dɪʔ| gᵚθ‖bə ˌdɪswɜ̃n ˈwɪpt‖]

This girl produces a frontal lisp for *guess* (transcribed as [θ]) and labializes the /ɹ/ of "*r*ipped." She also substitutes voiced [d] and [ʤ] sounds for the /t/ and /ʧ/ targets in "*t*ea*ch*er."

Examining childhood apraxia of speech

Childhood apraxia of speech (CAS) is a motor speech disorder. Children with (CAS) have difficulty planning and producing the movements of the articulators needed for intelligible speech, but muscle weakness or paralysis

doesn't cause it. In this sense, it is a *praxis* (planned movement) disorder, similar to adult AOS, which I discuss in "Transcribing Apraxia of Speech (AOS)" earlier in this chapter. However, because CAS affects children, it has a different cause and involves divergent symptoms, depending on the child's age and severity. For more information, see `www.asha.org/public/speech/disorders/childhoodapraxia.htm`.

A child with CAS will typically sound choppy, monotonous, or incorrect in stress placement. The unfamiliar listener will have difficulty understanding him/her. Longer words and phrases will be more difficult than shorter words. Speech may show *groping* (visible search behavior for sounds) and discoordination.

Here is a transcription of a 3-and-half-year-old American English-speaking girl diagnosed with CAS. She is talking about a playground.

> "It doesn't have a swing."
>
> [|ɪʔ dɛʔ hæ̃(C̲̄) ə hĩŋ‖]

The ExtIPA symbol (C̲̄) indicates an indeterminate consonant. This transcription suggests severely impaired consonant production, excess nasalization, and glottal stop substititions.

Part V
The Part of Tens

"I got a PhD in phonetics mainly because it just sounded so good."

Enjoy an additional Phonetics Part of Tens chapter online at www.dummies.com/extras/phonetics.

In this part . . .

- Identify and avoid ten mistakes that beginning transcribers often make.
- Figure out how you can improve your transcriptions and make fewer errors.
- Examine ten myths about English accents so you don't embarrass yourself when discussing them.

Chapter 20

Ten Common Mistakes That Beginning Phoneticians Make and How to Avoid Them

In This Chapter

- Knowing how to handle vowels
- Keeping track of stressed and unstressed syllables
- Getting your consonants correct
- Dealing with "r" quality in vowels and consonants

This chapter takes a closer look at ten common errors that newbie phoneticians can make when studying the International Phonetic Alphabet (IPA) and transcription. I give some pointers about what you can do to avoid making these common pitfalls.

Distinguishing between /ɑ/ and /ɔ/

Many newer phonetics students have difficulty telling the difference between the vowels /ɑ/ and /ɔ/. They're the hardest to distinguish because many North American dialects are merging these two back vowels.

To help you keep track of these two vowels, keep these hints in mind:

- To produce the /ɑ/, the mouth is more open; it's a low vowel with the jaw and tongue placed in the relatively lowest position. To produce the /ɔ/, the tongue and jaw are somewhat higher up, and the lips are usually somewhat rounded.
- If you *must* think of spelling (I don't generally recommend it; rely on what you hear), /ɔ/ is more commonly spelled "aw" or "ough" and a common spelling of /ɑ/ is "o" as in "h*o*t."

- /ɑ/ is typical in most American English productions of "f*a*ther," "h*o*spital," and "psychology."
- /ɔ/ is typical in most American English productions of "l*aw*," "c*ough*," and "s*ore*."

Refer to Chapter 7 for more information about these two vowels.

Getting Used to /ɪ/ for -ing spelled words

The vowel /ɪ/, which is a front mid-high lax vowel and International Phonetic Alphabet (IPA) small capital I, is a phonetic compromise case because this vowel changes its quality in a noticeable way in certain settings.

Before most *-ing* endings, people really don't produce fully tense front vowels, in productions like "run*eeeng*," because doing so would sound odd. On the other hand, most American speakers don't ordinarily say "runnin'" (/ˈɹʌnɪn/) in a formal setting, either. In reality, people usually produce a compromise case of "i" that is in-between an /i/ and an /ɪ/, a situation that phoneticians describe as neutralization before a nasal. Phoneticians use the lax character, /ɪ/, for these cases.

To avoid using /i/ by mistake, just remember that spelling does not work for the "i" in *-ing* endings. This is a case where small cap I (/ɪ/) takes over.

For the word "running," it's /ˈɹʌnɪŋ/. Notice also that the *-ing* ending can sometimes be pronounced with a "hard g" (IPA /g/), and sometimes not.

Staying Consistent When Marking /ɪ/ and /i/ in Unstressed Syllables

Most American talkers don't produce a fully tense /i/ at the end of a word, such as "ready," which sounds like "read*eeee,*" nor a completely lax /ɪ/, as in a Southern-accent "read-*ih*." Instead, the vowel is a compromise — it's somewhere between a tense /i/ and a lax /ɪ/. For this reason, some phoneticians transcribe an unstressed syllable as the tense member of the pair, such as /ˈɹɛdi/ while other phoneticians transcribe it as /ˈɹɛdɪ/. In this book, I use the tense "i" ending, [i].

To avoid confusion, decide on one transcription system and stick with it. That way, you can account for any regional variation you hear.

Knowing Your R-Coloring

The IPA rules for *rhoticization,* also called *r-coloring,* can seem a bit maddening, and many phonetics students commonly have problems remembering when r-coloring is indicated by having a vowel followed by an "r" such as in /ɑɹ/, /ɪɹ/, and /ɔɹ/ or when the IPA vowel characters themselves are marked for rhoticization with a special diacritic. For some reason, the crazy rules give the mid-central vowels special privilege. Chapter 2 discusses the mid-central vowels, which have the "uh" (/ʌ/ and /ə/) and "er" (/ɜ˞/ and /ɚ/) sounds. These vowels (and only these vowels) have their "r"-ness marked with a *diacritic*, a helper mark to further refine the meaning of an IPA character. This diacritic is a little squiggle placed on the upper right-hand side.

The remaining vowels may also have r-coloring, but it's indicated in the IPA by having an "r" consonant placed after them. Chapter 7 discusses the English vowels with their common pronunciations in American and British English.

Using Upside-Down /ɹ/ Instead of the Trilled /r/

This tip applies mainly to work with English, because the alveolar trill, /r/, is used in many world languages, including Afrikaans, Spanish, and Swedish. The English /ɹ/ is generally described as either a bunched or apical approximant and is represented in the IPA as /ɹ/.

To avoid using the wrong "r" when transcribing, keep these exercises in mind:

- **Practice producing alveolar trills.** Let the tip of your tongue move in the airstream as you say some words in other languages, such as the word for "mule" in Spanish, "burro" (/ˈburo/) or the word for "step" in Polish, "krok" (/ˈkrɔk/).
- **Read and separately contrast phonemes.** Focus on these: /r/, /ɹ/, /ara/, /aɹa/, /roː/, /ɹoː/. Remember, the /ː/ at the end of a vowel means extra long.

Check out Chapter 7 for more details about the upside-down /ɹ/ and trilled /r/.

Handling the Stressed and Unstressed Mid-Central Vowels

Some beginners have trouble knowing when to use an /ʌ/ versus an /ə/, or an /ɜ˞/ versus an /ɚ/. Many beginning transcribers mix up these mid-central

vowel characters. Just remember that you find both plain and r-colored schwas in English in unstressed syllable positions. That is, both schwa, /ə/, (as in "th*e*" or "*a*ppear") and "schwar", /ɚ/, (as in "teach*er*" and "p*er*formance") are in unstressed syllables. The other two mid-central vowels occur in stressed syllables, such as "D*ou*g" and "c*ur*tain."

Forming Correct Stop-Glide Combinations

As a beginning transcriber, you'll face many stop-glide combos that can cause you potential troubles. *Glides* are the consonants /j/ and /w/, so-called because they are vowel-like but don't form the core (nucleus) of a syllable. They're a *natural class* (they're a meaningful grouping) of the English approximants. Here are a couple combinations that you need to know how to form:

- *Palatalized* are stop consonant-palatal combinations where the palatal approximant has an immediate effect on the sound of the stop. Thus, you can easily distinguish the minimal pair "coot" versus "cute" — /kut/ versus /kjut/.
- *Labialized* are stop consonant-labiovelar combinations where the approximate also affects the stop, as in "kite" versus "quite" — /kaɪt/ versus /kwaɪt/.

To avoid making these types of mistakes (such as calling a "cutie" a "cootie"), refer to Chapter 6 where I provide more tips to help you.

Remembering When to Use Light-l and Dark-l

The alveolar lateral approximant consonant (IPA /l/) in English has two *allophones*. The two are as follows:

- **Light l:** When /l/ is produced at the beginning of a syllable, it's generally articulated with the tongue tip or blade near the alveolar ridge. Doing so gives it a higher sound, a "light l." You transcribe this allophone as /l/. Try it! Say "*l*a *l*a *l*a!" Don't you feel lighter already?

  Think of the word "*l*ight" starting with an "l"; this is the "light l" in the syllable-initial position (/laɪt/ in IPA).
- **Dark l:** The "dark l" is produced in the velar region. Think of the word "dorsa*l*" (/ˈdɔɹsəɬ/). You write this allophone as [ɬ] in IPA. Say "full," "pal," and "tool," and you should be able to feel your tongue rise in the rear of the oral cavity.

Remember these two "l" allophones in this way: Light "l" will never occur before consonants or before a pause, only before vowels. However, dark "l" doesn't occur before vowels.

Transcribing the English Tense Vowels as Single Phonemes or Diphthongs

Sometimes you just have to make up your mind. For the English sounds in the words "b*ai*t," "b*ee*t," "b*oa*t," and "b*oo*t," you can represent the vowel qualities in at least two different ways. At a basic level, these sounds can be described as simple monophthongs /e/, /i/, /o/, and /u/. More accurately, these English tense vowels have *offglides* (a changing sound quality toward the end) and are therefore better described as diphthongs: /eɪ/, /ij/, /oʊ/, and /uw/. Many phoneticians follow the conventions used in this book and apply this mixed set of symbols: /eɪ/, /i/, /oʊ/, and /u/.

To avoid making mistakes, decide which system to use and stick with it. Refer to Chapter 7 for more information on English vowels.

Differentiating between Glottal-Stop and Tap

Newbie transcribers also often have trouble telling the difference between the glottal stop and the voiced alveolar tap, which are two quite different gestures. Here is a quick overview of the two.

- **Glottal stop:** It takes place deep in the throat and can literally kill you if you hold it for too long. Its IPA symbol looks like a question mark without the dot: [ʔ].
- **Voiced alveolar tap:** It's an innocent little tap in your mouth that marks you as a quintessential American or Canadian. Its IPA symbol looks like a small pawn chess piece: [ɾ].

  The thing that they have in common is they're both allophones for alveolar stops in English; they can both stand in for a /t/ or /d/.

Refer to Chapter 6 where I provide more information about each so you can avoid using them incorrectly.

Chapter 21

Debunking Ten Myths about Various English Accents

In This Chapter

- Figuring out the different American accents
- Eyeing British accents
- Looking at the Australians and Canadians

A rich accent inventory comes with numerous varieties of English spoken throughout the world. Many people hold negative beliefs about certain dialects or accents of English for no other reason than "they sound funny." This chapter shows how some speakers' common assumptions about English accents, in reality, have little or no linguistic basis. This chapter debunks some common myths people have about different English accents.

Some People Have Unaccented English

One common myth is that some people are fortunate not to have accents. In fact, everyone has an accent. Even different members in the same family may have slightly different versions of the same regional accent. To *dialectologists* (a linguist or phonetician who specifically studies dialects), each person's individual accent is called an *idiolect,* which is an individual variant of a dialect. Dialects vary based on where you live, who you hang out with as a kid, what schools you attended, what TV shows you watched, your personality, and yes, your family. Therefore, because everyone is slightly different, determining who the people are who don't have accents makes no sense.

Two important points are relevant, according to the field of *sociolinguistics* (the study of language and language use in society), to explain that everyone has accents:

- Speakers of a language frequently make judgments of language *prestige* (which language is preferred or sounds the best) preference, with positive preference tilting toward the upper classes and negative preference against the lower classes. Note this is different than saying that someone has no accent.
- Accent judgments are subjective. For some people, English accents that traditionally are viewed negatively (such as Cockney English or African American English) can be cool!

At a practical level, something about this idea of everyone having an accent is clearly true. For instance, in North America English, accents and grammars that are markedly different from GAE or are difficult to interpret can be an impediment to one's advancement in the corporate world of business, education, and finance. For this reason, many speech language pathologists work with accent reduction as a part of their practice. The goal of this specialty is to help individuals reduce foreign or regional accents to improve intelligibility so that clients may better adapt to their work and social situations.

Yankees Are Fast-Talkin' and Southerners Are Slow Paced

When I moved to Dallas, the mailman greeted me and asked me if I was a Yankee. I told him I wasn't. This seemed to give him some relief.

"Ya know," he said, "I just can't stand 'em. When there's someone pushin' on your back in the market, it's a Yank. Rush, rush, rush! All the dang time!"

This struck me as the flip side of the insulting stereotypes about the "slow, stupid Southerners" common in so many movies and TV shows. You have to wonder if these fast/slow generalizations are true at least with respect to speech.

A number of recent studies actually do provide evidence that a geographic dialect factor influences speaking rate. Professors at Ohio State University recently found that a group of Northern speakers (from Wisconsin) spoke significantly faster than a group of Southern speakers (from North Carolina). This type of finding has been reported in previous studies, including regional dialect differences observed within other countries (England and Holland).

These studies don't say anything about niceness, smartness, or the tendency to push people in the back in the supermarket. However, for better or worse, some people may assume that these behavorial characteristics coincide with articulation rate. There is reason to believe that along with acquiring a regional dialect people might assume a certain articulation rate.

British English Is More Sophisticated Than American English

Some people think that British English is better or classier than American English. However, nothing is more sophisticated about British or American English (and their many dialects). They're simply different. Be careful to distinguish between the perfectly natural response of enjoying the sound and feel of various accents from deciding that a certain accent means a particular language (or group of language users) is sophisticated or not.

For example, assuming that one monolithic dialect known as British English compared to American English exists isn't realistic. Which accents are actually being considered? British Received Pronunciation (RP)? Estuary English? Cockney? Many people in the United States and Canada tend to equate British speech (specifically RP) with positive prestige, which means they look at British English as having a higher social value.

In most cases young countries that have descended from older ones often view the older country's accent with prestige. One notable exception is Portugal and Brazil, where Brazilian Portuguese is apparently the preferred form, and European Portuguese speakers now aspire to sound more like Brazilian Portuguese speakers.

Minnesotans Have Their Own Weird Accent

Speakers in Minnesota speak a variety of dialects, predominantly North Central American English. Parts of Montana, North Dakota, South Dakota, Minnesota, regions of Wisconsin and Iowa, and Michigan's Upper Peninsula share this dialect.

A Minnesotan may sound exotic to a Texan or somebody from York, England, but no more so than somebody from Wisconsin or Upper Michigan. The media may have perpetuated the idea that Minnesotans have something particularly

odd going on with their speech; however, because Minnesotans share this dialect with their neighbors, nothing is particular or peculiar about speech in Minnesota.

American English Is Taking Over Other English Accents around the World

Another myth suggests that American English is dominating the other English accents around the world and slowly taking them over, yet little evidence actually suggests this takeover. People learning English as a second language (ESL) are often interested in both American- and British-accented English, say in a country such as Japan. The ESL industry is booming in the United Kingdom and shows no indication of being colonized by predatory North Americans.

English has many wonderful varieties, which I discuss in Chapter 18, including Irish (Hibernian), New Zealand, Australian, South African, and Indian. Most of the citizens of these countries are doing quite well with their English dialects and don't have a burning need to replace them with the American brand.

People from the New York Area Pronounce New Jersey "New Joysey"

Although some speakers from this area (and, by the way, from New Orleans) produce mid-central r-colored vowels in a different fashion than ordinary speakers of GAE, it doesn't reach a so-called "oy," that is /ɔɪ/.

Instead, these talkers produce a more subtle off-glide, more like /eɪ/. In the New York City area, very few talkers actually have this accent. Movies have probably preserved the memory of this urban legend.

British English Is Older Than American English

Saying that British English has been around longer than American English isn't necessarily true, especially depending on which British English you're talking about. Just because English originated in England, which means the *roots* of English are more British than American, doesn't mean all British English is older.

Languages are always changing and many words and formations in British English today are likely just as new (or perhaps newer) than comparable American words. This phenomenon is also true with dialects. Compared to some of the newer British dialects (such as Estuary English), many American dialects are ancient.

The Strong Sun, Pollen, and Bugs Affected Australian English's Start

Some people still must believe that Australian English began because the early Australians had to close their mouths because of the sun, pollen, and bugs. Actually, present-day Australia started out as the colony of New South Wales, in 1788. The native-born children were exposed to a wide range of different dialects from all over the British Isles, including Ireland and South East England. Together, this generation created a new dialect.

A controversy surrounding Australian dialects today concerns the basis for variation. Most phoneticians maintain that there is relatively little geographical variation in Australian dialects and that Australian English primarily reflects individual social status. Others suggest subtle and detectable regional differences may exist.

Canadians Pronounce "Out" and "About" Weirdly

Canadian raising is the raising of the core of the two English diphthongs (/aɪ/ and /aʊ/) so that their core vowel (/a/) is replaced by a more central vowel, such as /ʌ/:

/aɪ/ → [ʌɪ]

/aʊ/ → [ʌʊ]

This sound change is a well-known characteristic of many varieties of Canadian English. To make a raised Canadian diphthong, say "house" beginning on a mid-vowel core, /hʌʊs/. Nor are Canadians saying "aboot" or "aboat" for "about." However, to someone unfamiliar with the dialect, it may sound like that. Non-Canadians may hear a somewhat exaggerated pronunciation of these vowels. This is because the diphthong is starting from a different position in the vowel space.

Although Canada has become famous for this sound change, it's also quite common in New England, including the regional accent of Martha's Vineyard, as well as parts of the upper Midwest. How abeut that?

Everyone Can Speak a Standard American English

Modern phonetics is descriptive, not prescriptive, which means that phonetics seeks to describe the sounds of the world's languages, not to make policy recommendations. For this reason, any general tendencies are referred to as GAE, not Standard. After all, if your speech is standard, what does that make mine? Substandard?

Such judgments are perhaps interesting, but they're the stuff of sociolinguistics and social stratification theory — not phonetics.

Most phoneticians apply norms that decide what GAE is. Phoneticians use these norms, for instance, to distinguish GAE pronunciation of the word "orange" /ɔɹnʤ/ from non-American accents (such as Scottish, /ˈaɾɪnʤ/), or regional American accents (New York City /ˈɑɹɪnʤ/). However, these GAE definitions are nevertheless quite broad. Someone on the West Coast of the United States would be within the bounds of GAE when he pronounced "orange" as a single-syllable word/ˈɔɹnʤ/, as would someone on the East Coast when she pronounced the word bi-syllabic but with a mid-high vowel instead of back low in vowel initial position /ˈɔɹɪnʤ/.

Index

• Symbols and Numerics •

• A •

• B •

• C •

• D •

• E •

• F •

• G •

• H •

• I •

• J •

• K •

• L •

• M •

• N •

• O •

• *p* •

• Q •

• R •

• S •

• T •

• U •

• V •

• W •

• X •

• Y •

• Z •

About the Author

William F. Katz, PhD, graduated from Brown University in 1986 and is a professor of communication sciences and disorders at the University of Texas at Dallas. He teaches phonetics, speech science, and aphasiology. His research focuses on neurolinguistics, including the breakdown of speech and language in adult aphasia and apraxia. He has developed novel techniques for correcting speech errors based on visual feedback of articulatory movement. These techniques are designed to help adults with communication disorders subsequent to brain damage and second language learners who are working on accent reduction.

Dedication

To my dearest wife, Bettina, for months of having to hear about *the book* again. And to teenagers Hannah and Sarah for putting up with dad's eternally stupid jokes.

Author's Acknowledgments

Thanks to all who helped make this work possible. To Janna, Paul, and Kayla for great suggestions from the get-go, Titus, Jackie, Linus, Rivka *(Movie Star),* Mathias and the girls, Benji, and the rest of the Swiss crew for their support from afar. June Levitt contributed some wonderful ideas from her own teaching experiences. Sonya Naya Mehta helped with much of the graphics. Wiley editors Anam Ahmed and Chad Sievers patiently sculpted my raw enthusiasm into an actual tome that someone just might consider reading. Murray Munro, superb linguist and phonetician, helped wrinkle out some of the more egregious technical and scientific *faux pas*, although any remaining boo-boos are certainly not to be pinned on him, but on me, myself, and I. Profound gratitude goes to my teachers for introducing me to this wonderful field. Finally, thanks to the many phonetics students here in Texas whose plentiful questions have kept me on my feet.

> *"Much have I learned from my teachers, more from my colleagues, but most from my students."*
>
> — Talmud: Ta'anit, 7a-1, R. Hanina

Publisher's Acknowledgments

Associate Acquisitions Editor: Anam Ahmed

Project Editor: Chad R. Sievers

Copy Editor: Chad R. Sievers

Technical Editors: Murray Munro, PhD, and Sibley Slinkard

Senior Project Coordinator: Kristie Rees

Cover Photos: © BSIP SA/Alamy

Apple & Mac

iPad For Dummies,
5th Edition
978-1-118-49823-1

iPhone 5 For Dummies,
6th Edition
978-1-118-35201-4

MacBook For Dummies,
4th Edition
978-1-118-20920-2

OS X Mountain Lion
For Dummies
978-1-118-39418-2

Blogging & Social Media

Facebook For Dummies,
4th Edition
978-1-118-09562-1

Mom Blogging
For Dummies
978-1-118-03843-7

Pinterest For Dummies
978-1-118-32800-2

WordPress For Dummies,
5th Edition
978-1-118-38318-6

Business

Commodities For Dummies,
2nd Edition
978-1-118-01687-9

Investing For Dummies,
6th Edition
978-0-470-90545-6

Personal Finance
For Dummies,
7th Edition
978-1-118-11785-9

QuickBooks 2013
For Dummies
978-1-118-35641-8

Small Business Marketing Kit
For Dummies,
3rd Edition
978-1-118-31183-7

Careers

Job Interviews
For Dummies,
4th Edition
978-1-118-11290-8

Job Searching with
Social Media
For Dummies
978-0-470-93072-4

Personal Branding
For Dummies
978-1-118-11792-7

Resumes For Dummies,
6th Edition
978-0-470-87361-8

Success as a Mediator
For Dummies
978-1-118-07862-4

Diet & Nutrition

Belly Fat Diet For Dummies
978-1-118-34585-6

Eating Clean For Dummies
978-1-118-00013-7

Nutrition For Dummies,
5th Edition
978-0-470-93231-5

Digital Photography

Digital Photography
For Dummies,
7th Edition
978-1-118-09203-3

Digital SLR Cameras &
Photography For Dummies,
4th Edition
978-1-118-14489-3

Photoshop Elements 11
For Dummies
978-1-118-40821-6

Gardening

Herb Gardening
For Dummies,
2nd Edition
978-0-470-61778-6

Vegetable Gardening
For Dummies,
2nd Edition
978-0-470-49870-5

Health

Anti-Inflammation Diet
For Dummies
978-1-118-02381-5

Diabetes For Dummies,
3rd Edition
978-0-470-27086-8

Living Paleo For Dummies
978-1-118-29405-5

Hobbies

Beekeeping
For Dummies
978-0-470-43065-1

eBay For Dummies,
7th Edition
978-1-118-09806-6

Raising Chickens
For Dummies
978-0-470-46544-8

Wine For Dummies,
5th Edition
978-1-118-28872-6

Writing Young Adult Fiction
For Dummies
978-0-470-94954-2

**Language &
Foreign Language**

500 Spanish Verbs
For Dummies
978-1-118-02382-2

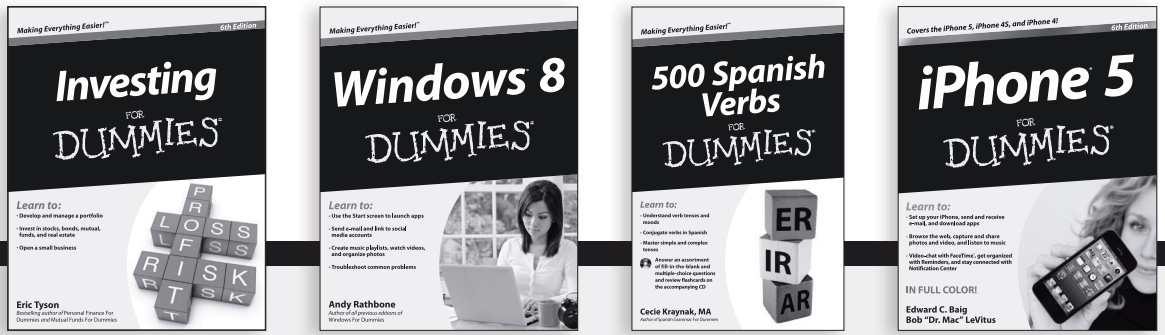
English Grammar
For Dummies,
2nd Edition
978-0-470-54664-2

French All-in One
For Dummies
978-1-118-22815-9

German Essentials
For Dummies
978-1-118-18422-6

Italian For Dummies
2nd Edition
978-1-118-00465-4

Available in print and e-book formats.



Available wherever books are sold. For more information or to order direct: U.S. customers visit www.Dummies.com or call 1-877-762-2974. U.K. customers visit www.Wileyeurope.com or call (0) 1243 843291. Canadian customers visit www.Wiley.ca or call 1-800-567-4797.

Connect with us online at www.facebook.com/fordummies or @fordummies

Math & Science

Algebra I For Dummies, 2nd Edition
978-0-470-55964-2

Anatomy and Physiology For Dummies, 2nd Edition
978-0-470-92326-9

Astronomy For Dummies, 3rd Edition
978-1-118-37697-3

Biology For Dummies, 2nd Edition
978-0-470-59875-7

Chemistry For Dummies, 2nd Edition
978-1-1180-0730-3

Pre-Algebra Essentials For Dummies
978-0-470-61838-7

Microsoft Office

Excel 2013 For Dummies
978-1-118-51012-4

Office 2013 All-in-One For Dummies
978-1-118-51636-2

PowerPoint 2013 For Dummies
978-1-118-50253-2

Word 2013 For Dummies
978-1-118-49123-2

Music

Blues Harmonica For Dummies
978-1-118-25269-7

Guitar For Dummies, 3rd Edition
978-1-118-11554-1

iPod & iTunes For Dummies, 10th Edition
978-1-118-50864-0

Programming

Android Application Development For Dummies, 2nd Edition
978-1-118-38710-8

iOS 6 Application Development For Dummies
978-1-118-50880-0

Java For Dummies, 5th Edition
978-0-470-37173-2

Religion & Inspiration

The Bible For Dummies
978-0-7645-5296-0

Buddhism For Dummies, 2nd Edition
978-1-118-02379-2

Catholicism For Dummies, 2nd Edition
978-1-118-07778-8

Self-Help & Relationships

Bipolar Disorder For Dummies, 2nd Edition
978-1-118-33882-7

Meditation For Dummies, 3rd Edition
978-1-118-29144-3

Seniors

Computers For Seniors For Dummies, 3rd Edition
978-1-118-11553-4

iPad For Seniors For Dummies, 5th Edition
978-1-118-49708-1

Social Security For Dummies
978-1-118-20573-0

Smartphones & Tablets

Android Phones For Dummies
978-1-118-16952-0

Kindle Fire HD For Dummies
978-1-118-42223-6

NOOK HD For Dummies, Portable Edition
978-1-118-39498-4

Surface For Dummies
978-1-118-49634-3

Test Prep

ACT For Dummies, 5th Edition
978-1-118-01259-8

ASVAB For Dummies, 3rd Edition
978-0-470-63760-9

GRE For Dummies, 7th Edition
978-0-470-88921-3

Officer Candidate Tests, For Dummies
978-0-470-59876-4

Physician's Assistant Exam For Dummies
978-1-118-11556-5

Series 7 Exam For Dummies
978-0-470-09932-2

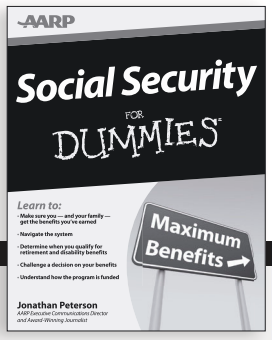
Windows 8

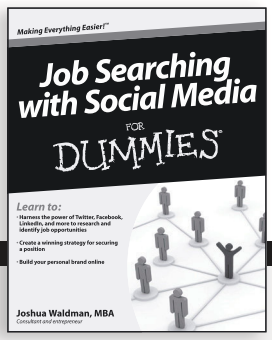
Windows 8 For Dummies
978-1-118-13461-0

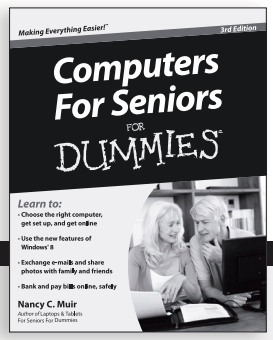
Windows 8 For Dummies, Book + DVD Bundle
978-1-118-27167-4

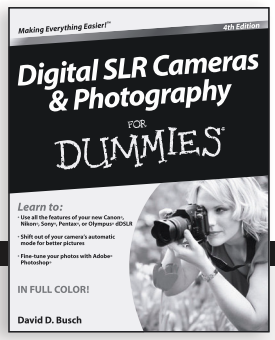
Windows 8 All-in-One For Dummies
978-1-118-11920-4

Available in print and e-book formats.









Available wherever books are sold. For more information or to order direct: U.S. customers visit www.Dummies.com or call 1-877-762-2974. U.K. customers visit www.Wileyeurope.com or call (0) 1243 843291. Canadian customers visit www.Wiley.ca or call 1-800-567-4797.

Connect with us online at www.facebook.com/fordummies or @fordummies

Take Dummies with you everywhere you go!

Whether you're excited about e-books, want more from the web, must have your mobile apps, or swept up in social media, Dummies makes everything easier .

Visit Us

Like Us

Follow Us

Watch Us

Join Us

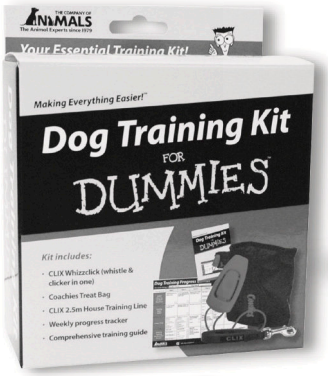
Pin Us

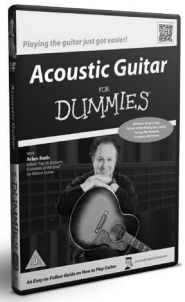
Circle Us

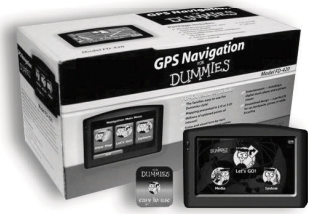
Shop Us

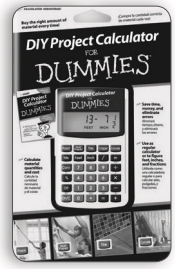
Dummies products make life easier!

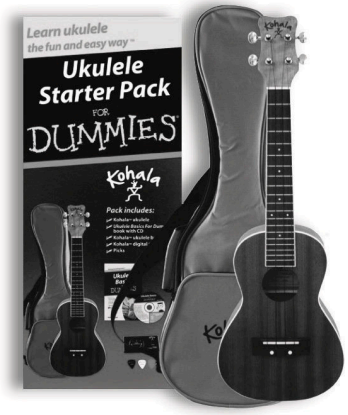
- DIY
- Consumer Electronics
- Crafts
- Software
- Cookware
- Hobbies
- Videos
- Music
- Games
- and More!

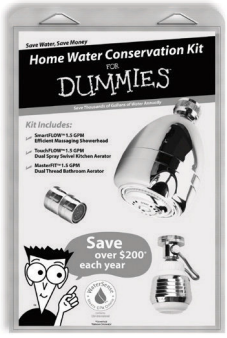












For more information, go to **Dummies.com®** and search the store by category.

FOR DUMMIES®
A Wiley Brand

Mobile Apps FOR DUMMIES®

There's a Dummies App for This and That

With more than 200 million books in print and over 1,600 unique titles, Dummies is a global leader in how-to information. Now you can get the same great Dummies information in an App. With topics such as Wine, Spanish, Digital Photography, Certification, and more, you'll have instant access to the topics you need to know in a format you can trust.

To get information on all our Dummies apps, visit the following:

www.Dummies.com/go/mobile from your computer.

www.Dummies.com/go/iphone/apps from your phone.