1 Introducing Phonetics: The Science of Speech

Speech is the most fundamental means of human communication. Nearly all of us—with the exception of the profoundly deaf and people with severe cognitive dysfunction—begin learning to speak during the early stages of childhood and continue to use spoken language as a mode of interaction. It is mainly through speech that we establish and develop our most important personal attachments, acquire the cultural competence that allows us to function as members of society, and pass on our wisdom to our offspring.

An especially fascinating aspect of speech is that it appears to be a uniquely human capability. Other social animals engage with their communities in a variety of ways, and we humans often talk affectionately to our canine and feline companions. But non-human animals don't use anything quite like speech with one another; nor do they carry on spoken interactions with us. The importance of speech as a social tool and its status as a defining characteristic of humanness have made phonetics a dynamic and fruitful area of study since ancient times. In fact, the phonetician John Ohala (2006) suggests that phonetics may be the oldest of the behavioral sciences and, in many respects, one of the most successful. Its value is evident in the wide range of useful things we can do because of speech-related research and technological innovations. You might not think of the telephone, for instance, as a product of phonetics, but its invention has revolutionized our lives because Alexander Graham Bell devoted his life to the study of speech. Even more impressive are today's artificial talking agents, which can read web pages aloud for people with visual disabilities and provide a voice to those who, like the late Stephen Hawking, have lost the ability to speak. The reverse situation is also becoming a practical reality in the form of computer applications that accept speech input and follow instructions to help us complete daily tasks. It is less well known to the public that criminal investigations sometimes employ forensic phonetics during the collection of evidence leading to the conviction of

Applying Phonetics: Speech Science in Everyday Life, First Edition. Murray J. Munro.

Companion website: www.wiley.com/go/munro

offenders. On the medical front, advances in the study of speech anatomy and physiology have led to effective remediation for many types of vocal tract injuries, language delays, and speech disorders. And with respect to language preservation and revitalization, linguists are able to apply phonetic principles along with speech technology to systematically document the sound patterns of endangered languages.

1.1 speech, language, and communication

1.1.1 classifying communication types

When phoneticians talk about speech, they mean the component of language conveyed by sound. Speech is rooted in human biology in that it is produced through the centrally mediated (i.e., managed by the brain) activity of the vocal tract. Of course, many animals can make non-speech sounds, so to understand what makes speech special, we need to consider its relation to other communication types. Figure 1.1 shows some of the ways that different types of communication may be classified. Broadly, *communication* refers to an exchange of information between organisms. The first thing to notice is that some communicative behaviors count as linguistic and some do not. This distinction is shown on the horizontal dimension of the figure. *Language* is an elaborate symbolic system that can be used to convey all sorts of information from one person to another. But many kinds of information can certainly be conveyed without it. Animals often send messages, for example, using body postures and movements, cries and roars, and even odors. Linguists would generally agree that none of these forms of expression counts as language. Humans, too, can convey a great deal of useful information without language. Babies express emotional states like discomfort, frustration, and pleasure through cries, giggles, and other vocalizations, and through facial expressions that are not linguistic either. Most of the time, classifying something as language or not is straightforward, yet a

	NON-LINGUISTIC	LINGUISTIC
NON-VOCAL	bee "dance" cat scent marking babies' facial expressions adults' body language	writing sign languages
VOCAL	baby cries, coos, giggles adult sighs, screams, throat clearing frog croaks, cat meows, dog barks parrot "talk"	human speech
SYNTHETIC	facial expressions, postures, and movements of animated characters	artificially-generated speech

Figure 1.1 Classification of communication types

fully satisfactory technical definition of *language* has proved surprisingly elusive. We will return to this issue shortly.

A second dimension of communication—shown on the vertical dimension of Figure 1.1—concerns whether or not a vocal tract is involved. Across animal species, many forms of *non-vocal* communication are possible. Some of these appear to be simple, as when a dog leaves its signature by urinating on a fire hydrant or a cat rubs the side of its head against a piece of furniture, leaving a scent marking. These signs are primitive in that they transmit relatively little information ("I've been here!") and are not directed at any particular receiver. Other examples of non-vocal communication, however, are information-rich and geared to a specific, interested audience. Bees perform an elaborate "dance" in view of the other members of their hive, using a sophisticated non-vocal system that informs the community of both the location and the quality of a food source, and does so with a high degree of precision.

Human non-vocal communication varies in its complexity as well. A gentle touch with a hand can be understood as an affectionate act, and facial expressions can reveal a wide range of emotional states. More impressively, human non-vocal behavior is sometimes linguistic. This is because human language has more than one means of transmission or MODALITY. The vocal modality is the one we call speech, but the written modality is used for books, computer documents, emails, and text messages. While the written and vocal modalities have a great deal in common, they also differ in important respects. For instance, when we talk, we usually don't use the same level of formality and the same vocabulary as we do when writing a business memorandum. And when we write, we are not able to take advantage of certain aspects of spoken language that affect how our ideas are conveyed, such as changes in tempo, loudness, or pitch.

In addition to the vocal and written forms of language, another modality is *gesture*, a means by which well over 100 distinct languages, including *American Sign Language* (**ASL**), *Japanese Sign Language* (Nihon Suhwa), and *Spanish Sign Language* (Lengua de Signos Española), are transmitted. Sign systems are exceptional among human languages in that they do not use sound at all; however, they are every bit as complex and nuanced as spoken languages.

Vocal communication, shown in the middle row of Figure 1.1, entails the production of sound using parts of the body that are also used for breathing and eating. While many animals, including humans, can vocalize, not all their sounds fall into this category. Crickets chirp or *stridulate* by rubbing parts of their wings against each other, and humming birds hum because of rapid movements of their wings. But virtually all tetrapods have a LARYNX, a structure in the upper part of the body that serves a variety of purposes and happens to make vocalization possible. The meow of a cat, the bleating of a goat, and even the hissing of a snake are all the result of exploiting laryngeal structures, together with other parts of the VOCAL TRACT, to create non-linguistic sounds. When a baby screams out in frustration because it is hungry, it is communicating vocally but non-linguistically; so too are adults when they sigh, gasp, or clear their throats to attract attention.

Speech, however, is more complicated than these non-linguistic sounds because it is an expression of what we consider *true* language. But how do we

Tetrapods are creatures that evolved from four-footed ancestors. Note that tetrapods themselves do not necessarily have four feet. In fact, snakes and birds are tetrapods because their evolutionary predecessors were four-footed reptiles. And all mammals, including humans, are tetrapods as well.

define *language*? Doing so in a succinct way has turned out to be extremely difficult, so linguists have sometimes preferred to focus on certain properties, termed *design features* by Charles F. Hockett, which, taken together, might capture the difference between language and other communicative systems (Hockett & Hockett, 1960). While we won't go into all of his original 13 features here, four of them that are especially relevant to speech are ARBITRARINESS, DISCRETENESS, PRODUCTIVITY, and DUALITY OF PATTERNING. Human speech is arbitrary in the sense that there is generally no connection between the things that are referred to and the spoken symbols used to represent them, as is true for the words *tree* (/tɹi/), *moon* (/mun/), and *love* (/lʌv/). A non-English-speaker hearing these words for the first time would not be able to guess their meanings from the way they sound. In fact, two different languages sometimes assign the same sequence of sounds to entirely different meanings. For instance, the Japanese word for *tree* happens to be /ki/, pronounced like the English word *key*.

Discreteness refers to our interpretation of the speech signal as a sequence of individual segments, which makes it possible to structure words and other linguistic units in terms of smaller chunks. These PHONES are familiar to us as vowels and consonants, and they occur in every language. Next, *productivity* accounts for our ability to arrange these phones in countless different orders to convey distinct meanings. A simple example is our ability to analyze the word *cat* as a series of three phones represented as /k/, /æ/, and /t/ in the INTERNATIONAL PHONETIC ALPHABET. Note that we can arrange the same sounds in two other orders that have distinct meanings:

- /t/ + /æ/ + /k/ gives us the sequence /tæk/, spelled as *tack*. (Don't be misled by the spelling!)
- /æ/ + /k/ + /t/ gives us the sequence /ækt/, spelled *act*.

Linguists use an asterisk (*) to denote a non-existent word or impossible sound sequence in a particular language.

The last of the four features, *duality of patterning*, is closely related to discreteness and productivity and refers to the way spoken languages make use of a system of sounds that relate to a system of meanings. While the individual phones of a language are typically meaningless on their own, they can be combined in orderly ways for communicative purposes. When we say that they work as a system, we mean that within a particular language there are restrictions on how they can be combined. In English, for instance, we can't have */tkæ/, */ætk/, or any other words starting or ending with /tk/. However, we *can* use an additional property of our language to systematically change meanings: we add /s/ to *cat*, *tack*, and *act* to create the plural form of each. The four phones we've mentioned occur in thousands of other words. In fact, the remarkable consequences of discreteness, productivity, and duality of patterning become clear when we realize that English has only about 35–38 phones (depending on the dialect), which can be combined to create a vocabulary of perhaps a million words. What's more, it is perfectly possible to use new combinations of phones to invent new words. As far as I know, *splenk* (/splɛŋk/) is not an English word, but there is nothing to stop an inventor from developing a new household tool and calling it a *splenk*. And splenk users would realize, without being told, that the plural form of the word is likely *splenks*!

It is quite easy to pinpoint some of the ways in which non-human communication systems lack the properties of human language. For instance, the bee dance mentioned earlier expresses the angle of the sun relative to a food source through the angle at which the dance is performed. In that case, there *is* a connection between the communicative symbol and the thing it stands for, so in that respect bee communication lacks arbitrariness. In a similar vein, there is no evidence of discreteness or productivity in the meowing of a cat; it isn't possible to break a cat vocalization down into smaller pieces and rearrange them to generate new messages.

However, we must not make the mistake of being too categorical in our assessments of human versus non-human communication systems. In the first place, animal vocalizations are *not* entirely devoid of speech-like elements. Vervet monkeys, for example, use different vocal alarm calls to alert other members of their group to imminent dangers. Seyfarth, Cheney, and Marler (1980) studied these vocalizations by recording them, analyzing their acoustic composition, and playing the sounds back through loudspeakers while observing the vervets' behavior. The monkeys produced a low-pitched, grunt-like call, for instance, on the approach of an eagle, and a higher-frequency call at the sight of a snake. When other vervets were exposed to only the recorded calls with no visual stimuli, they responded as if an eagle or snake were present. On the one hand, the alarms apparently aren't divisible into smaller units, and there is no indication that vervets can rearrange the sounds of one call to create another call with a different meaning. Consequently, the vervet vocalizations can be said to lack discreteness and productivity. But, on the other hand, the calls do have the speech-like property of arbitrariness: there appears to be no connection between the sounds and the things they represent. In that respect, they share something with human speech.

A second observation is that human speech does not fully conform to the design features we've mentioned. For one thing, not all aspects of speech are arbitrary. English, and presumably all other languages, has scores of *onomatopoeic* words like *bang*, *burp*, *chirp*, and *clap*, which sound to some degree like the things they refer to. Research has also uncovered intriguing examples of SOUND SYMBOLISM, in which particular speech sounds are associated with certain meanings (Westbury, Hollis, Sidhu, & Pexman, 2018). For instance, in linguistic judgment tasks, people tend to link the sounds /k/ and /t/ to the concept of *sharpness*, while /m/ and /l/ suggest *roundness*. These and other non-arbitrary mappings may be much more than trivial matters. Some evidence indicates that they may facilitate child language acquisition. We will return to this topic when we discuss its applicability in the complexities of product naming in Chapter 14.

1.1.2 technology and our changing understanding of "speech"

Several decades ago, we would have stopped with the four-way classification of communication types that we have developed so far. But contemporary technology is changing our understanding of the nature of communication. Suppose that an animated character in a movie or a video game displays facial expressions and gestures indicating anger or a threat of violence. Of course,

the character itself has no feelings or desire to communicate, but the animator has created a representation of non-linguistic, non-vocal communication that is readily grasped by viewers. In a similar vein, computers do not volitionally use speech with an intent to communicate (at least, not at present!). However, we have no problem calling artificially-created utterances *speech*, even though they can be generated entirely without a vocal tract. Such utterances have a communicative function, whether the purpose is to give voice to a human user who cannot speak, to "read" a text aloud to a blind person, or to convey an account balance to a bank customer over the phone. We can capture these recent developments by adding a third category to the vertical dimension of the grid to cover *synthetic* communication types, both non-linguistic and linguistic.

To sum things up so far, *communication* refers to the transmission of a message from one organism or entity to another; *language* is a means of communication that uses arbitrary symbols; and *speech* consists of communicative sounds produced in the vocal tract or synthetically.

1.2 the sound structure of speech

The APSSEL website provides a link to where you can download Praat and some instructions on getting started.

Figure 1.2 is a visual representation of an English sentence ("The museum hires musicians every evening"), which was generated from speech using an application called *Praat* (Boersma & Weenink, 2019). Depictions like this are of great use in phonetics research, and we will discuss them in more detail in later chapters. For now, it is enough to know that the top portion is an ACOUSTIC WAVEFORM capturing the oscillations of air particles when a speaker utters something into a microphone. The lower panel is a SPECTROGRAM illustrating the sound frequency components of speech, with lower frequencies at the bottom of the display. Dark regions in the spectrogram indicate concentrations of acoustic energy. Notice that the utterance appears as a variable acoustic pattern with occasional abrupt changes in darkness and shape. However,

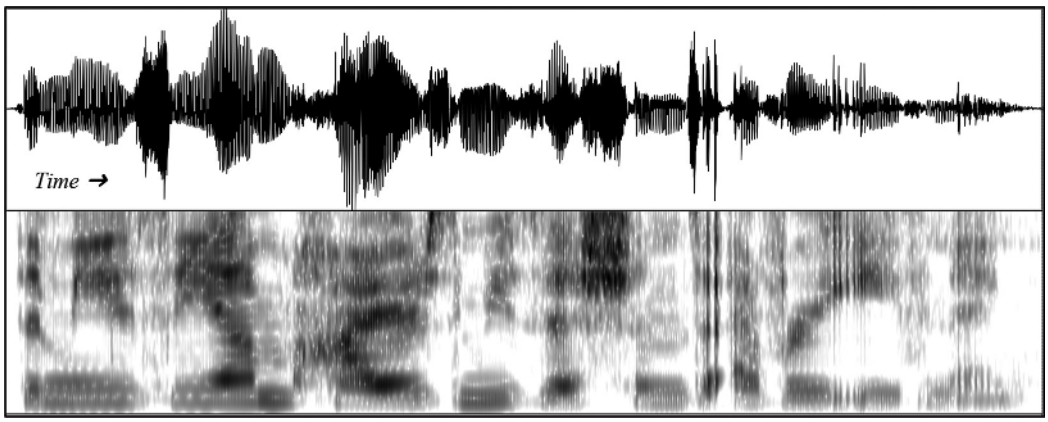

Figure 1.2 Acoustic waveform (top) and spectrogram (bottom) of "The museum hires musicians every evening," as produced by an adult female speaker

finding discreteness (as described earlier) in this representation turns out to be quite a challenge. In some instances, it is a straightforward matter to locate the beginning and end of a word, but in others it is much more difficult. Often, it is not possible to find clear demarcations between individual vowels and consonants within the words because these units overlap one another to varying degrees. What this means is that the phenomenon of discreteness is actually not an aspect of the acoustic signal itself. Rather, it is something that we humans partially impose upon the speech stream we hear. Put another way, *discreteness* is the result of the way we interpret vocally-produced sound. This apparent lack of a one-to-one relationship between sound and perception adds an interesting layer of complexity to our understanding of the nature of speech—one that we will revisit throughout this book.

TRY THIS ☛ Download and install the Praat software on your computer, and record yourself saying the sentence in Figure 1.2. Use the software to display a waveform and spectrogram as in the figure. Compare your own production with the one in the figure.

1.3 phonetics as a field of study

Phonetics focuses on the sounds of language rather than on written forms. Moreover, phoneticians generally accept the primacy of speech over the written modality. Historically, it must have preceded writing because many world languages have a spoken form but no written one, yet we know of no natural language that can be written but not spoken. Another reason for assuming the primacy of speech is that children become highly proficient in oral language well before they are capable of reading and writing. In fact, many people never learn to read or write at all. No one seriously doubts that literacy is an important aspect of human culture, but it is a mistake to regard written language as *more* important or more "correct" than speech. That simply isn't true. Writing was invented by humans as a way of representing the spoken word on the printed page, and not the reverse. Nor is it true that printed texts are good models of how we should speak. It *is* true, however, that written style in many languages differs from spoken style. The formal usage in English academic textbooks is quite different from everyday spoken English, while English speech and writing are much more similar in other situations such as text messaging. For that reason, we cannot regard writing and speaking as different manifestations of the same thing. Sometimes they are very close; other times they are not.

1.3.1 branches of phonetics

Despite its concentration on sound, phonetics is a diverse field with a multidisciplinary reach—so much so that you will have to read this entire book to get a good picture of its scope. To begin, let's consider its three core branches, as shown in Figure 1.3.

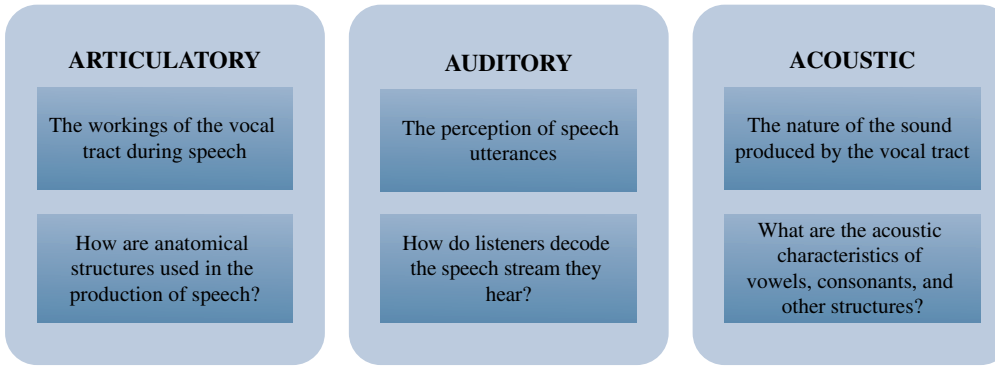

Figure 1.3 Core branches of phonetics

Articulatory phonetics covers the anatomy and physiology of speech, with a focus on the structures we use to generate vocal sounds, such as the tongue, lips, nose, and larynx, and on the ways in which these anatomical components function together. An understanding of normal articulation is essential for helping people with speech disorders, who may need the assistance of a speech–language pathologist. A child who has undergone surgery for a cleft palate, for instance, may experience production difficulties that can be remedied by a speech professional. The domain of *auditory phonetics* is the structures and processes through which the human auditory system decodes the speech stream into meaningful messages. Numerous insights into child language acquisition have been gained because of perceptual research within this branch of the field. Finally, *acoustic phonetics* addresses the physical properties of the speech sounds themselves, often through analyses like the ones depicted in Figure 1.2. Thanks to acoustic-phonetic research, we are able to synthesize the intelligible, natural-sounding speech now available on computers, phones, and other household devices.

It will help you to remember the core areas of phonetics if you think of them as the three As: Articulatory, Auditory, and Acoustic.

Though the core branches provide us with one way of appreciating the nature of the field, another set of descriptors can be used to characterize phoneticians' *approaches* to their work. Historically, phonetics has relied extensively on the trained human ear, and even today, careful listening and skillful transcription are fundamental to IMPRESSIONISTIC phonetics. Some highly skilled phoneticians have played a key role in criminal cases by providing ear-based analyses of threatening phone calls. However, thanks to the technological advances of the twentieth century, INSTRUMENTAL phonetics has taken a more prominent role than ever before. It involves the use of a variety of sophisticated tools for imaging the vocal tract during speech and for pinpointing important acoustic details. A third approach, the most recently developed, which we will refer to as AUTOMATIC SPEECH PROCESSING, uses artificial intelligence (**AI**) for a variety of purposes, including computer recognition of speech, forensic voice identification, and speech synthesis.

Finally, APPLIED phonetics, which is heavily emphasized throughout this book, is concerned with the ways in which our understanding of speech can be used to achieve practical ends. Strictly speaking, applied phonetics is not

separate from any of the types we've mentioned already. In fact, it makes use of research findings from all three core branches and can be approached impressionistically, instrumentally, and through AI. One of the first applications of phonetics to come to mind for most people is language teaching. In fact, the speech sciences have been powerfully influenced by ideas about spoken language instruction, as we'll see in Chapter 11. However, there are many other sub-branches of applied work, including forensic phonetics and clinical phonetics, as well as specialized areas of application relating to accessibility, automation, music, animation, stage and screen acting, and business.

Cat-to-Human: Feed Me!

Our family cat, Owen, happens to be extraordinarily vocal when he is hungry, needs to play, or spots a squirrel through the window. The left panel of Figure 1.4 shows an acoustic representation of one of his productions, which you can compare with the much more complex, human-produced word on the right. While it's tempting to attribute specific meanings to the sounds he makes, the available evidence does not allow researchers to say what exactly cats "mean" when they vocalize. Nonetheless, a wealth of studies of domestic feline behavior have helped shed light on the problem. It is well established, for instance, that cats vocalize differently when they are around humans than they do in feral conditions. Perhaps the most famous study of feline communication is Mildred Moelk's "Vocalizing in the House-Cat: A Phonetic and Functional Study," published in 1944 in the *American Journal of Psychology*. Moelk identifies three kinds of production: mouth closed, mouth gradually closing, and mouth tensely open. The resulting sounds fall into 16 different patterns, for which she even provides transcriptions using the International Phonetic Alphabet. According to Moelk, the patterns correspond to (among other things) greetings, acknowledgments, demands, begging, and bewilderment.

Several perceptual experiments have shown that humans can accurately interpret cat vocalizations, at least to some extent. Schötz and van de Weijer (2014) collected audio recordings of cats in two contexts: at feeding time and while waiting at the veterinarian. They played the meows in random order to 30 human listeners tasked with guessing the context of each token. Not only did the listeners perform correctly at above chance levels, but those having experience with cats as pets scored higher than those without. So, it seems that humans can learn to partially interpret feline messages through experience. In another study by McComb, Taylor, Wilson, and Charlton (2009), humans judged solicitation purrs (recorded during food-seeking) as more urgent and less pleasant than non-solicitation purrs played at equal volume. The listeners' ability to make the distinction, which occurred with or without cat experience, was due to a particular high-frequency component within the solicitation purrs that has a surprising parallel in human infant cries. In the authors' words, the less pleasant purrs "may be exploiting an inherent mammalian sensitivity to acoustic cues relevant in the context of nurturing offspring" (p. R507). If so, then when Owen howls from the kitchen, it's not simply because he has learned what gets our attention. Rather, his biology may be calling out to our own!

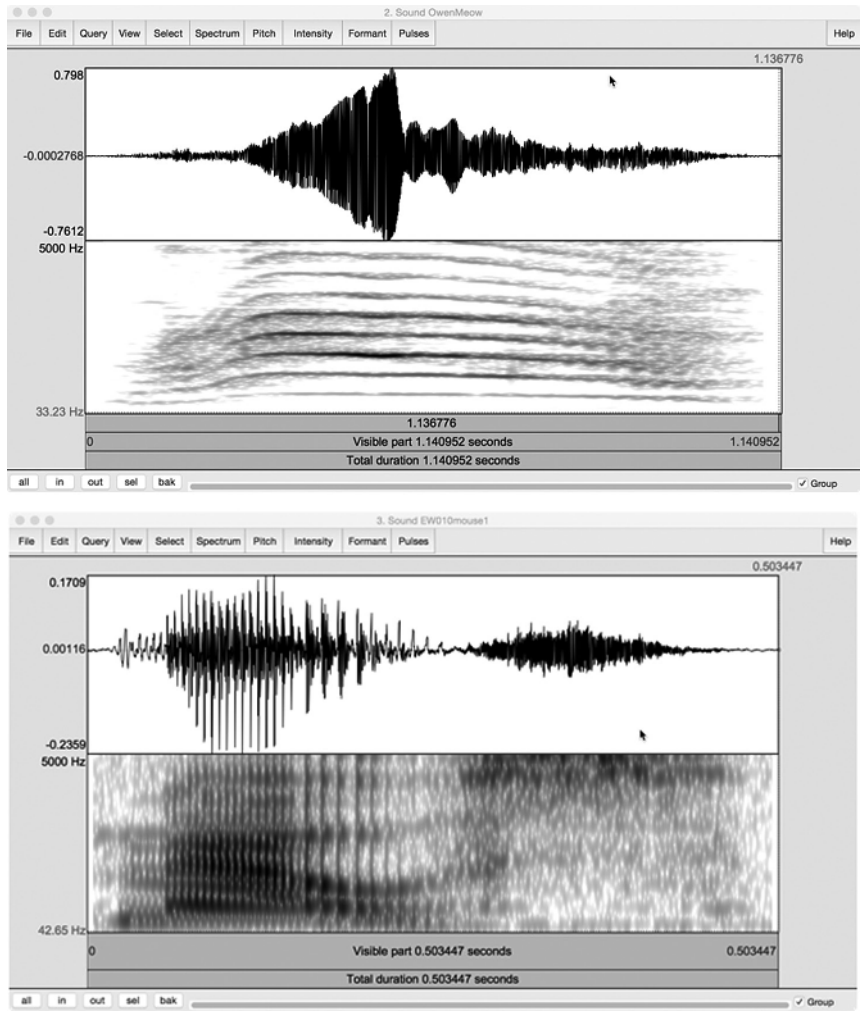

Figure 1.4 Waveforms and spectrographic representations of a cat's meow (top) and a human production of the word "mouse" (bottom)

for further thought, analysis, and discussion

1. Find an audio or video recording of a politician, entertainment figure, or sports personality responding extemporaneously (without preparation) to questions during an interview. Write out several of the speaker's utterances, including any hesitation forms like "um" or "ah" and any speech errors. Identify some of the ways in which the speaker's productions differ from written language.

2. With a classmate, discuss how many phones you think occur in each of the words below. Suggested answers are provided on the APSSEL website.

a) map	e) speech	i) cheese
b) shot	f) school	j) tick
c) fix	g) waited	k) often
d) wheat	h) marched	l) epitome

3. Record yourself producing the words in item 2 using Praat, and examine the waveforms and spectrograms. (Consult the APSSEL website for basic instructions.) First, consider the appearance of the vowel portions of each word, taking note of the repeating patterns. Next, compare these with the noisy sounds: the sound of "s" in *speech* and *school*. Now compare the sounds of the letters *m*, *w*, and *t*, which all appear in more than one word. How do they differ from one another? Does a particular sound look identical when it appears in different words?
4. While teaching reading, one of my grade-school teachers frequently told the class that we needed to pronounce the words the way they were spelled. "Look at the letters. Sound the words out!" she would exclaim, sometimes impatiently. What do you think she meant by this? Was she right?