1 Phonetics

1.1 Speech sounds

PHONETICS is the study of SPEECH SOUNDS, and we can divide it into three sub-fields. The study of the movements of the SPEECH ORGANS in SPEECH PRODUCTION is called ARTICULATORY PHONETICS. The study of the physical properties of speech sounds as sounds is called ACOUSTIC PHONETICS. Finally, the study of listener responses to speech sounds in SPEECH PERCEPTION is called AUDITORY PHONETICS.

Most of what you learn about phonetics in this book will be about articulatory phonetics. The main reason is that you can get a lot of information about the movements of your speech organs just by paying careful attention to the sensations those movements cause. Acoustic phonetics and auditory phonetics require machines and mathematics right from the start, but you can understand fairly sophisticated articulatory descriptions without getting into equipment and experiments. On the other hand, this emphasis on articulation doesn't mean that we'll completely ignore the acoustic and auditory domains, and experimental results will figure in the discussion frequently. You'll find a very basic introduction to acoustic displays in §1.12 below.

Since articulatory phonetics involves describing the activities of the speech organs, even a basic introduction like this one has to refer to certain aspects of human anatomy and physiology. Figure 1-1 is a schematic view of the VOCAL TRACT, which you can think of as a kind of branching passageway from the lungs to the lips and nostrils.

It's important to keep in mind that the speech organs all have other, more basic functions. For example, although the tongue plays a very important role in speech production, its role in eating is clearly primary. But even though the speech functions of the speech organs are secondary in this sense, these functions are probably more than just incidental. The reason for thinking

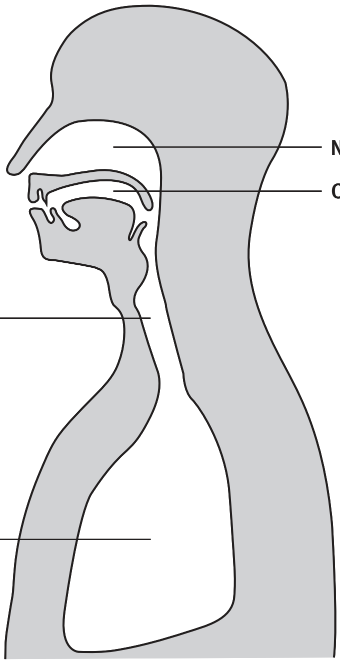

Figure 1-1 Vocal tract

so is that some features of speech-organ anatomy and physiology suggest adaptation for speech.[1] For example, the basic functions of the LARYNX are (1) to seal off the lungs and keep the chest rigid during physical exertion, and (2) to prevent foreign material from getting into the lungs. But the human larynx doesn't seem to be optimally adapted either for protecting the lungs or for maximizing respiratory efficiency. At the same time, it does seem to be adapted for producing speech sounds.[2]

Not all sounds that people produce with their speech organs are speech sounds. For instance, coughing and throat clearing aren't used as speech sounds in any language. Of course, there are also sounds that are speech sounds in some languages but not in others. For instance, CLICKS are used as consonants in some languages of Africa, but they aren't used as speech sounds in English or Japanese.[3] It's even possible for the same sound to be used sometimes as a speech sound and sometimes not. A famous illustration of this point involves comparing the sound of blowing out a candle with the sound represented by the *wh* at

[1] For the idea that the speech functions of the speech organs are only incidental, see Sapir 1921:8–9. On speech-organ adaptation, see Hockett 1960:127, Aitchison 1998:149–51, Hudson 2000:144–6. For a skeptical view, see Sampson 1997:24–5, 55–60.
[2] On the human larynx in speech and survival, see Lieberman 1991:53–7, 1998:47–8, 85–6.
[3] For details on clicks, see Ladefoged and Traill 1984, 1994. Textbook treatments include Laver 1994:173–9, Ladefoged and Maddieson 1996:246–80.

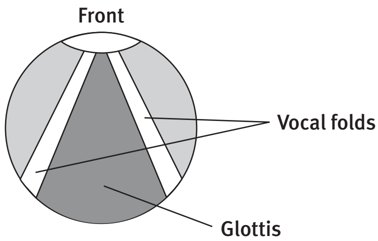

Figure 1-2 Vocal folds and glottis from above (adapted from Roca and Johnson 1999:15)

the beginning of an English word like *where* (assuming a variety of English in which *where* and *wear* are pronounced differently).[4] The two sounds may not be exactly identical, but they're at least very nearly the same, and the point of the example is clear: in terms of function, they're completely different.

1.2 Airstream mechanisms

To produce any sort of audible signal you have to produce sound waves by putting air into motion. An AIRSTREAM MECHANISM is a way of initiating such motion, and every speech sound is produced by one of four airstream mechanisms. The most important by far is the PULMONIC-EGRESSIVE airstream mechanism, in which the muscles of the chest cavity work like a bellows to expel air from the lungs.[5] All languages use the pulmonic-egressive airstream for the majority of their speech sounds, and many languages, including both Japanese and English, use only this airstream. In other words, all the consonants and vowels of Japanese and English are produced by various modifications of the pulmonic-egressive airstream on its way through the vocal tract. Incidentally, it isn't difficult to speak while breathing in rather than breathing out, but no language in the world uses this PULMONIC-INGRESSIVE airstream mechanism for ordinary speech.[6]

1.3 Phonation

The VOCAL CORDS or VOCAL FOLDS are two shelf-like structures in the larynx that can be held apart or brought together, and the space between the vocal folds is called the GLOTTIS. When the vocal folds are held apart the glottis is open, and when the vocal folds are held together the glottis is closed. In Figure 1-2 the glottis is open for normal breathing.

[4] Sapir 1925:37–9.
[5] Textbook treatments of airstream mechanisms include Catford 1988:19–32, Laver 1994:161–83.
[6] Ladefoged and Zeitoun 1993.

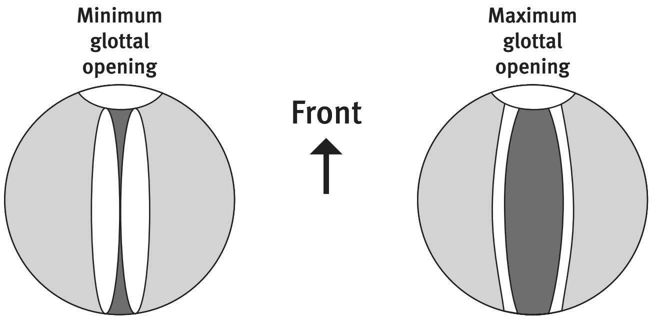

Figure 1-3 Extreme positions during vocal-fold vibration cycle (dapted from Roca and Johnson 1999:16)

If the vocal folds are held close together and there's sufficient airflow, they'll vibrate, that is, open and close rapidly again and again. This vibration is known as PHONATION or VOICING.[7] The two illustrations in Figure 1-3 show the most closed and most open positions that the vocal folds reach in the course of a vibration cycle.

Speech sounds accompanied by voicing are VOICED, and those not accompanied by voicing are VOICELESS. For example, the English words *zip* and *sip* begin with consonant sounds that are identical except that the former (spelled *z*) is voiced, whereas the latter (spelled *s*) is voiceless. Not only can you hear voicing, you can feel it. Place the fingers of one hand on the side of your neck as if feeling for your pulse. If you pronounce a long *zzzzz* sound, you can feel the vibration of your vocal folds through your skin. If you pronounce a long *sssss* instead, you won't feel the vibration, since this sound is voiceless.

1.4 Nasality

The rear portion of the roof of the mouth, called the SOFT PALATE or VELUM, can be moved, as shown in Figure 1-4. When the velum is open (lowered), as on the right, air flowing out of the lungs can escape through the nose. When the velum is closed (raised), as on the left, the ORAL CAVITY is sealed off from the NASAL CAVITY. Speech sounds produced with the velum open are NASAL, and those produced with the velum closed are ORAL. For example, the English words *mat* and *bat* begin with consonant sounds that are nearly identical, but the former (spelled with *m*) is nasal, whereas the latter (spelled with *b*) is oral.

[7] Textbook treatments of phonation include Catford 1977:93–116, Laver 1994:184–201.

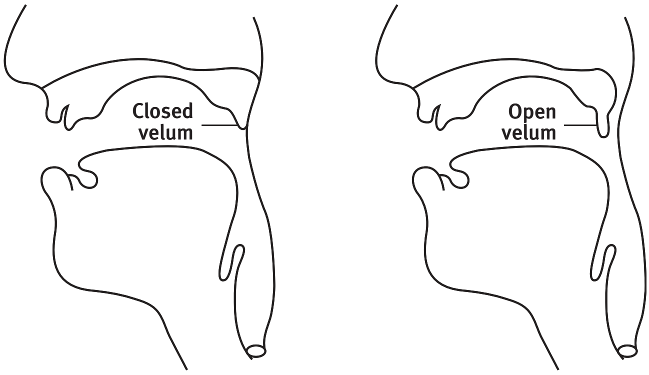

Figure 1-4 Velum and nasality (adapted from Ashby 1995:31)

1.5 Transcription and segments

A PHONETIC TRANSCRIPTION of a stretch of speech consists of a sequence of PHONETIC SYMBOLS. These symbols have fixed, conventional values that are independent of any particular language. The INTERNATIONAL PHONETIC ASSOCIATION endorses a set of symbols called the INTERNATIONAL PHONETIC ALPHABET, and the acronym IPA can stand either for the association or for its alphabet.[8] Most linguists follow the IPA recommendations fairly closely, but many non-IPA symbols are in common use.[9] It's standard practice to enclose phonetic transcriptions in square brackets. For example, the English word *bed* can be transcribed [bɛd].

An individual consonant or vowel sound is called a PHONETIC SEGMENT, and each symbol in a typical phonetic transcription stands for a segment. This kind of representation suggests that each segment is a brief but static configuration of the speech organs and that speech production involves changing from one state to another with instantaneous transitions between states. In fact, though, actual speech production is nowhere near so neat. The speech organs very seldom maintain a static configuration for any measurable length of time.[10] Another problem is that when two segments are adjacent, the movements necessary to produce them overlap. For example, the vowel [u] in English *soon* is pronounced with rounded lips, and the vowel [i] in English *seen* is pronounced with spread lips. If you pronounce these two words and pay careful attention to your lips, you'll notice that the [s] in *soon* is pronounced with rounded lips, whereas the [s] is *seen* is pronounced with spread lips. In other words, the lip position during the [s] in each word anticipates the position that's necessary

[8] International Phonetic Association 1999 is the standard guide to the current version of the International Phonetic Alphabet.
[9] Pullum and Ladusaw 1986. [10] Abercrombie 1967:38.

for the following vowel. This overlapping or smearing of adjacent segments is known as COARTICULATION, and coarticulation effects often extend beyond immediately adjacent segments.[11] One way to think about coarticulation is to regard the speech-organ configuration associated with a given segment as a TARGET CONFIGURATION that isn't necessarily attained in actual speech.

As a result of coarticulation, segments are squashed together into syllables, and it's syllables that seem to be the basic units of production and perception.[12] As we'll see in §6.1, it turns out to be very difficult to give a satisfactory definition of syllables, but they do seem to be intuitively natural units for ordinary speakers. On the other hand, the fact that most people can learn to use an alphabetic writing system without too much difficulty indicates that segment-sized units must have some sort of psychological reality as well.[13]

The important point for now is just that segment-by-segment phonetic transcriptions involve a significant degree of abstraction. But such transcriptions are very convenient, and we'll rely on them heavily in this book. Much of the remainder of this chapter is concerned with the articulatory description of segments and with the IPA symbols used to transcribe them. You should always keep in mind that no transcription, no matter how careful, is really complete. Minute differences are generally ignored, but the analyst's native language, experience with other languages, and phonetic training all influence which details are included and which are left out.[14] A relatively precise transcription that provides a lot of details is called a NARROW TRANSCRIPTION, and a relatively imprecise one is called a BROAD TRANSCRIPTION.[15]

1.6 Length

Although segments don't in general correspond to static configurations of the speech organs (§1.5), some segments clearly last longer than others. For most purposes it's sufficient to distinguish two degrees of duration: short and long. In IPA transcription, a colon-like LENGTH MARK [ː] is written after the appropriate symbol to indicate that a particular segment is long. For example, the first vowel in *kōto* コート 'coat' is longer than the first vowel in *koto* 琴

[11] Farnetani 1997, Hardcastle and Hewlitt 1999. [12] Abercrombie 1967:37, Lieberman 1977:120–1.
[13] Liberman (1996:1–43) argues that a specialized phonetic mode of perception has evolved in humans that allows effortless decoding of acoustic syllables into their component segments. For the idea that alphabetic letters represent intuitively real segment-sized units, see Saussure 1959:38–9, Fromkin and Rodman 1998:498. For objections, see Aronoff 1992, Daniels 1992, Faber 1992, Port 2007. See also Prince 1992, Fujimura 2000.
[14] Bloomfield 1933:84–5, Pike 1943:109–10.
[15] The terms "narrow" and "broad" go back to Henry Sweet, who used "broad" to mean "phonemic" (see Chapter 2 below). For discussion, see Matthews 2001:32–5, International Phonetic Association 1999:28, Rogers 2000:46–7.

'koto': [koːto] versus [koto]. We'll look at Japanese length distinctions very carefully in later chapters (§3.2, §§5.1–2, §§5.4–5).

1.7 Suprasegmentals

Certain features of speech are ordinarily represented in phonetic transcriptions as superimposed on segments rather than being integral parts of segments. Such features are usually called SUPRASEGMENTAL, and they're represented this way because linguists regard them as properties of units larger than single segments. For example, in the English word *mama* the first syllable has greater prominence or STRESS than the second, and the degree of stress is usually considered a property of a syllable as a whole rather than a property of any of the individual segments that make up a syllable. The IPA transcription for a stress difference is a vertical stroke before the more prominent syllable: [ˈmɑmɑ]. In terms of production, a syllable with a higher degree of stress is produced with more energy.[16]

Another feature that's ordinarily treated as suprasegmental is PITCH. All voiced sounds involve vocal-fold vibration (§1.3), and listeners perceive differences in the FREQUENCY of this vibration as differences in pitch. Within the limits of a speaker's voice range, the rate of vibration is under voluntary control and is determined mainly by the tension of the vocal folds.[17] When a particular pitch pattern is a property of a syllable, it's called a TONE; when it's a property of a unit such as a phrase, it's called INTONATION. Descriptions of Japanese usually say that it has PITCH ACCENT. Roughly speaking, the idea is that pitch-accent patterns are properties of words, and a word's inherent pitch pattern has to be integrated into the intonation pattern of any phrase that contains that word. We'll look at the details of the Japanese pitch-accent system in Chapter 7.

1.8 Vowels

Vowels are produced by positioning the speech organs so that the vocal tract is free of significant obstructions above the glottis. Our focus here is on what's called VOWEL QUALITY. If two people who speak the same language pronounce the same vowel, the two versions will probably be different in several respects. One person's vowel might be longer than the other person's, or it might be louder, or it might be on a higher pitch. Also, every individual has a distinctive voice quality, which is why you can recognize a familiar person's

[16] Catford 1977:84–5, 1988:32–5, Laver 1994:511–23. [17] Hardcastle 1976:83–7.

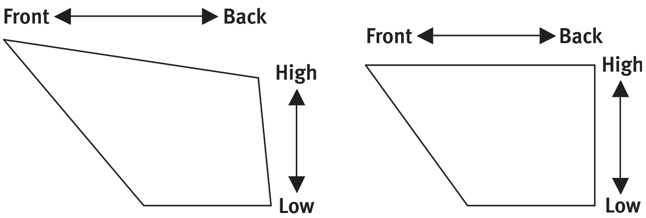

Figure 1-5 Stylized vowel-area diagrams

voice. Vowel quality is what's left when you ignore all these other differences. In other words, vowel quality is what makes a Japanese *o* sound an *o* as opposed to some other vowel.

The traditional description of a vowel articulation involves specifying the highest point of the tongue, with lip activity as an independent variable.[18] Many phoneticians have suggested that the targets for vowels are auditory and that articulatory descriptions are inadequate, but we'll consider typical vowel articulations in this section.[19] The entire range of possible positions for the highest point of the tongue is called the VOWEL AREA. In an x-ray image taken from the left side of a person's head, the lips are at the left, the back of the head is at the right, and the highest point of the tongue is somewhere in a two-dimensional space that corresponds to the vowel area. Phoneticians use a stylized diagram to represent the vowel area viewed from this angle, as in Figure 1.5. The version on the left in Figure 1-5 is closer to the actual shape of the vowel area, but the version on the right is a little easier to draw and much more popular.[20] All the vowel diagrams in the rest of this book will look like the version on the right.

We can specify the highest point of the tongue for any particular vowel in terms of the two dimensions of the vowel area: vertical (HIGH-MID-LOW) and horizontal (FRONT-CENTRAL-BACK). The distance between the highest point of the tongue and the roof of the mouth is larger in low vowels than in high vowels, and the jaw is more open, so you'll often see high vowels described as CLOSE and low vowels described as OPEN. The standard lip-position scale is SPREAD-NEUTRAL-ROUNDED, but unless extraordinary precision is called for, it's common practice to describe a vowel as simply rounded or UNROUNDED. On the other hand, as we'll see in §3.1, the spread-neutral-rounded scale is too crude to provide an adequate description of the Japanese vowel that's romanized as *u*.

[18] An additional parameter is necessary to specify the vowel qualities in some languages. See Ladefoged and Maddieson 1996:300–10.
[19] Ladefoged 1971:67–75, Lieberman 1977:138–54, Ladefoged 2007:166.
[20] Abercrombie 1967:157–9.

Table 1-1 American English vowels

Example word	Articulatory description	IPA symbol
beat	high front unrounded	[i]
bit	semi-high front unrounded	[ɪ]
bet	lower-mid front unrounded	[ɛ]
bat	semi-low front unrounded	[æ]
box	low back unrounded	[ɑ]
but	lower-mid back unrounded	[ʌ]
boss	lower-mid back rounded	[ɔ]
book	semi-high back rounded	[ʊ]
boot	high back rounded	[u]
baton	mid central unrounded	[ə]

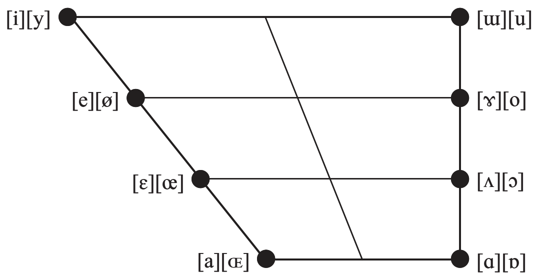

Figure 1-6 Cardinal vowels

The CARDINAL VOWELS are a set of reference vowels produced with tongue positions on the perimeter of the vowel area. The idea is that a person who knows the cardinal vowels can describe any possible vowel quite precisely by comparing it to the cardinal vowels.[21] Figure 1-6 shows the cardinal vowels displayed on a diagram of the vowel area. There are eight tongue positions, and for each tongue position, the symbol on the left specifies an unrounded vowel and the symbol on the right specifies a rounded vowel.

Table 1-1 shows how the vowels in some English words can be described using the vertical, horizontal, and lip-position dimensions. The table also includes the IPA symbols that are ordinarily used to transcribe these vowels. Of course, these descriptions apply only to a particular variety of English that I called UNITED STATES NEWSCASTER ENGLISH in the Preface. Throughout this book, the terms *English* and *American English* will refer to this variety unless I explicitly say otherwise. Figure 1-7 shows approximate tongue positions for these English vowels on a vowel-area diagram.

[21] Daniel Jones developed the system of cardinal vowels. For discussion, see Ladefoged 1967:50–142, Abercrombie 1967:151–62, Catford 1988:130–7, Laver 1994:273–5.

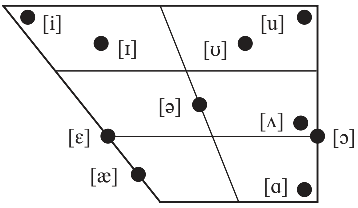

Figure 1-7 American English vowels

One noteworthy fact about English is that the only rounded vowels are back and non-low, and this same correlation between rounding and tongue position holds in many other languages as well. On the other hand, it isn't hard to find languages with front rounded vowels. For example, French has high front rounded [y] (as in *vu* [vy] 'seen') and higher-mid front rounded [ø] (as in *veut* [vø] 'wants'). It's also common for all the back non-low vowels in a language to be rounded, but as Table 1-1 shows, English has lower-mid back unrounded [ʌ] (as in *but*). Mandarin Chinese has higher-mid back unrounded [ɤ] (as in *hé* 河 'river'), and Korean has high back unrounded [ɯ] (as in *kul* 글 'writing'). We'll discuss whether or not Japanese has [ɯ] in §3.1.

In many languages vowel quality can change dramatically within a syllable. When the beginning and ending qualities are different enough to make it sound as if a syllable contains a sequence of two phonetically distinct vowels, the sequence is called a DIPHTHONG. A vowel that seems to maintain a steady quality throughout its duration, like those we considered above, is called a MONOPHTHONG. A diphthong is transcribed by writing appropriate vowel symbols for its beginning and ending points and connecting the two symbols with a ligature underneath. Table 1-2 shows how the diphthongs in some English words can be represented. Here again, the descriptions apply only to United States newscaster English.

Figure 1-8 shows these English diphthongs on a vowel-area diagram, using arrows to represent the movements in tongue position. In [o͜i] and [ɑ͜u], of course, the lip position changes along with the tongue position.

Vowels are ordinarily voiced, but VOICELESS VOWELS do occur in some languages, including Japanese. As we'll see in detail in §8.1 Japanese high vowels are often voiceless when surrounded by voiceless consonants. A voiceless vowel is transcribed by writing a small circle under the appropriate vowel symbol, so a typical pronunciation of *kita* 北 'north' could be represented as [ki̥tɑ]. Vowels can also be nasal, and such NASALIZATION is transcribed by writing a TILDE over the appropriate vowel symbol. For example, French *beau* [bo] 'beautiful' and *bon* [bõ] 'good' are identical except that the vowel in *beau* is oral, while the vowel in *bon* is nasal.