2 Segmental Articulation

This chapter gives a broad account of how speech sounds are made. After a brief introduction (2.1) the chapter begins with a functional overview of how speech is produced (2.2), a simple description of the various parts of the body used in speech (2.3) and some comments on the way in which we describe speech sounds (2.4).

The chapter then examines the means of producing a flow of air (2.5) and the role of the larynx in creating speech sounds (2.6). Subsequent sections turn more particularly to the articulatory nature of vowels (or vowel-like sounds, 2.7 and 2.8) and consonants (2.9).

Various aspects of articulation, concentrating on consonants, are then dealt with:

- the various places in the vocal tract at which consonants are made (2.10)
- the role of tongue position (2.11)
- different manners of consonant articulation (2.12)
- the shaping of constrictions (2.13)
- relative force of articulation (2.14)
- length (2.15)
- the timing of voicing (2.16).

2.1 Introduction

The human vocal apparatus is capable of producing a great variety of noises. Many of these do not count as speech sounds, such as coughs and snores and grunts, but we caution readers against being too narrow in their notion of speech sounds. It would be quite wrong to assume that English, or even Western European languages, are fully representative of phonological possibilities, and the range of sounds which we shall cover is far wider than occurs in any one language. In particular there are sounds, such as the kind of click sound which many English speakers use to express regret or disapproval (sometimes written as *tut* or *tsk*), which Europeans may well assume are not speech sounds, but which do occur in some languages.

In this chapter we will work towards developing a repertoire of all possible speech sounds and a framework in which to describe them – although, as we shall shortly see, we will do better to think in terms of human ability to make distinctions or differences in sound, rather than in terms of an inventory of sounds. To this end, we shall examine the function of the vocal apparatus as a speech-producing mechanism, and in the process show how it can be used to make all kinds of sounds.

2.2 A functional overview of the speech production process

We begin with a general functional overview of the process of speech production, but its more technical aspects are dealt with in detail in chapter 6. The human vocal apparatus can be viewed as a kind of mechanism – it has measurable dimensions, such as the distance from the larynx to the lips, it has moving parts such as the tongue, and so on. Figure 2.2.1 gives a simple functional model of this mechanism which omits almost all anatomical detail, but should help the reader through the outline description of the following paragraphs.

To produce sound of any kind, a source of energy is needed. For speech, a flow of air makes it possible to generate sounds, and the volume and pressure of the air supply determine the duration and loudness of sound produced. The majority of speech sounds (in fact *all* in English and Western European languages) use airflow from the lungs for this purpose. As shown in figure 2.2.1, the respiratory system therefore counts as the energy source, and the lungs form an air reservoir. The lungs are compressed by various respiratory forces, rather like a set of old-fashioned fire bellows. As the lungs are compressed, air flows out, and it is the periodic interruption, constriction and blockage of this airflow which results in the more or less continuous flow of sound which we identify as a sequence of speech sounds.

The airflow can be interrupted periodically by the vocal folds, which are situated in the airway above the lungs and form part of the air valve structure of the larynx. When airflow from the lungs through the windpipe is blocked by the closed vocal folds, air pressure below them builds up. This pressure momentarily forces the folds apart. As the air then flows out through the folds, the local air pressure is reduced and the folds can close again. Air is thus released in short puffs at a periodic rate. This process of vocal fold vibration, known as PHONATION, is similar to the process that produces noise when you inflate a balloon, stretch the neck into a thin aperture, and allow the air to escape through it. The puffs of air created by the vibration of the vocal folds occur at a certain rate or frequency. This frequency is variable and is determined by muscle forces controlling the tension of the vocal folds and by the

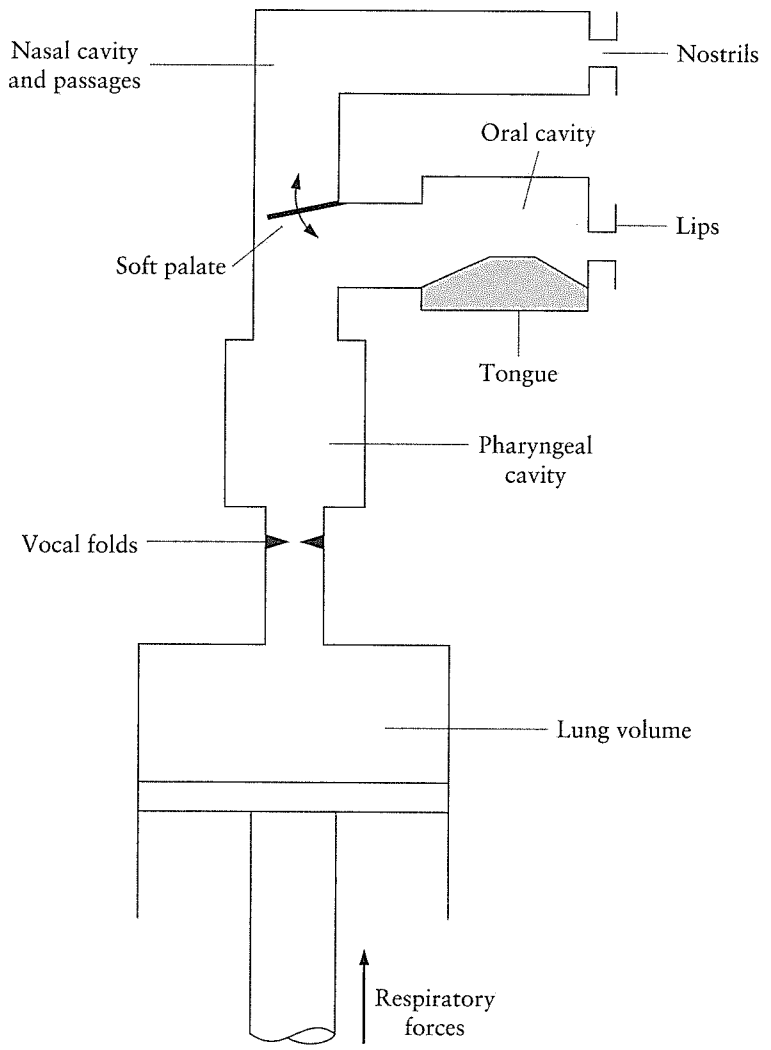

Figure 2.2.1 Functional model of the vocal tract

air pressure below the folds. The frequency is perceived as the PITCH of the voice. Sounds which are produced in this way, with air flowing from the lungs through vibrating vocal folds, include all vowels and vowel-like sounds.

These puffs of air constitute an effective sound source but are not in themselves sufficient to produce identifiable speech sounds. The essential additional ingredient is the contribution of the cavities above the vocal folds. These cavities can be opened or closed off and their size and shape can be manipulated in ways that modify the basic sound source, yielding a variety of individually identifiable speech sounds.

A simple example of this process is provided by the three vowel sounds heard in a typical southern English pronunciation of the words *heed*, *hard* and *hoard*.

(Appropriate phonetic symbols for the three vowels are [i:], [a:] and [ɔ:], where the colon is the convention for marking these vowels as relatively long.) In these vowels, the vocal folds vibrate as just described, releasing a periodic train of air puffs. The soft palate (which in normal quiet breathing hangs down to allow free airflow through the nasal cavity and nostrils) is raised, as it usually is during speech to stop or reduce airflow into the nasal cavity. Airflow therefore passes through the throat (pharyngeal cavity) and mouth (oral cavity). The shape, and hence the resonant properties, of these two cavities are controlled by the position of the tongue, the degree of jaw opening, and the shape of the lips. Thus for [i:] in *heed* the tongue is pushed forward and raised in the region just below the hard palate, while the lips are spread. For [a:] in *hard* the tongue is in a relatively neutral or slightly retracted position on the floor of the mouth, the jaw is opened further than for [i:], and the lips are opened in a neutral or natural position. For [ɔ:] in *hoard*, the tongue is retracted, the hump formed in it is raised some way towards the soft palate, and the lips are rounded. Each of these three articulatory positions alters the geometry of the pharyngeal and mouth cavities, and each position has its own characteristic resonant properties. The sound produced by the air puffs from the vibrating vocal folds is modified by these resonant properties, with the result that each vowel sound has a distinctive sound quality. Readers should be able to feel something of the change in articulatory setting if they say each of these three vowel sounds while paying attention to the position of the tongue, jaw and lips; it is also possible to verify the resonant effects of a cavity by producing an [a:] vowel while cupping and uncupping the hands around the lips.

Vowels and vowel-like sounds are made by varying the geometry of the pharyngeal and mouth cavities, but without any major obstruction or impediment to airflow. Consonantal sounds, on the other hand, are generally made by exploiting the articulatory capabilities of the tongue, teeth and lips in such a way that airflow through the mouth cavity is radically constricted or even temporarily blocked.

The [b] of the word *barn*, for example, is known as a STOP, produced as the name implies by transient blockage of the airflow. In this sound, the soft palate is raised to prevent airflow through the nasal cavity, the lips are closed for a fraction of a second, and, during this closure, air pressure builds up in the pharyngeal and mouth cavities. The lips are then parted, releasing the pressure behind them and allowing normal airflow for the vowel which follows. The characteristic sound of this articulatory action is largely due to the rapid changes in the resonant properties of the mouth cavity during the very short interval of time from the point when the lips begin to open to the point when normal vowel articulation has begun.

Other consonantal sounds rely on radical constriction of airflow within the mouth cavity, rather than transient blockage. Thus the [l] in the word *learn* is produced by holding the tip of the tongue against the ridge of flesh immediately behind the front teeth, and allowing airflow to be diverted around one or both sides of the tongue. Such sounds are known as LATERALS. This articulatory configuration again has its own particular resonant properties producing a characteristic quality of sound.

All sounds mentioned so far have relied on airflow through the pharyngeal and oral cavities. It is possible to block the oral cavity, so that air flows through the pharyngeal and nasal cavities, as in the [m] of the word *more*. Such NASAL consonants are produced with the soft palate lowered to allow airflow through the nasal passage, and with the mouth cavity blocked for the duration of the consonant. In this configuration, the unobstructed pharyngeal and nasal cavities and the blocked mouth cavity all contribute to the resonant properties of the sound.

Yet another way of producing consonantal sounds is by setting the articulatory organs in such a way that friction or turbulence is created. The simplest example of a FRICATIVE consonant is [h] as in *hard*. In this fricative, turbulence occurs both at the opening of the vocal folds and throughout the remainder of the airways and cavities through which air flows. In most fricatives, however, the sound is generated by air turbulence at some specific point. Thus the [v] in the word *vine* is produced with the lower lip held lightly against the edge of the upper front teeth, so that turbulence occurs when air is forced through.

In most of the sounds we have mentioned so far, vibration of the vocal folds continues through the sound. All such sounds are called VOICED. But some sounds are VOICELESS: they employ the same kinds of articulatory configurations that we have described for voiced sounds but the airflow is uninterrupted, as the vocal folds are not vibrating. There are now no periodic puffs of air to act as a sound source, and the constriction or interference somewhere in the airways and cavities above the larynx becomes the sound source. In a voiceless fricative, such as [f] in *fine*, for example, the turbulence created when the lower lip is held lightly against the edge of the upper front teeth is the sound source. Thus fricatives can be voiceless or voiced, and voiceless [f] is the counterpart of voiced [v], which uses both vocal fold vibration and the turbulence produced by localized constriction. Compare the words *fine* and *vine*, in which the principal distinguishing feature is the voicing, or vocal fold vibration, during the production of the [v].

Stops are also voiceless if vocal fold vibration does not begin until after the start of the release of the blockage in the mouth cavity. The major distinction between the initial sounds in the words *pat* and *bat*, as pronounced by most native speakers of English, is that in the former, vocal fold vibration begins after the lips have begun to part, and in the latter, the vocal folds are already vibrating when the lips part. (In fact there is more than one simple way of distinguishing between voiced and voiceless stops, but we shall return to this later.)

We have given only the briefest summary of some of the major types of articulatory processes involved in speech production. In normal continuous speech some of these processes occur very rapidly, and may interact with each other as a result. The sound output can show rapid changes of quality, and this dynamic aspect of speech is also important in providing cues that allow listeners to recognize a coherent sequence of speech sounds. And of course normal adult users of language are also aided by their knowledge of their language and their consequent expectations about what are, and are not, likely and acceptable sound sequences forming normal utterances.

2.3 The organs of speech

The term ORGANS OF SPEECH refers to all those parts of the human body which are concerned in various ways with the production of speech. Most of them are only secondarily concerned with speech production – their primary functions are to do with eating, chewing and swallowing food, and respiration. Figure 2.3.1 shows a section through the body indicating the major organs which contribute to the speech process.

The organs of speech shown in figure 2.3.1, namely the lungs, trachea, larynx, the pharyngeal and oral cavities with their component parts, and the nasal passages, constitute as a group what is termed the VOCAL TRACT. For functional and descriptive purposes, the tract is normally divided into two basic parts, one above the larynx, the other below it. Within the larynx itself are the vocal folds: the aperture between the folds is known as the GLOTTIS, and the tract above the glottis is therefore called the SUPRAGLOTTAL vocal tract, and that below it the SUBGLOTTAL vocal tract. The choice of this point of

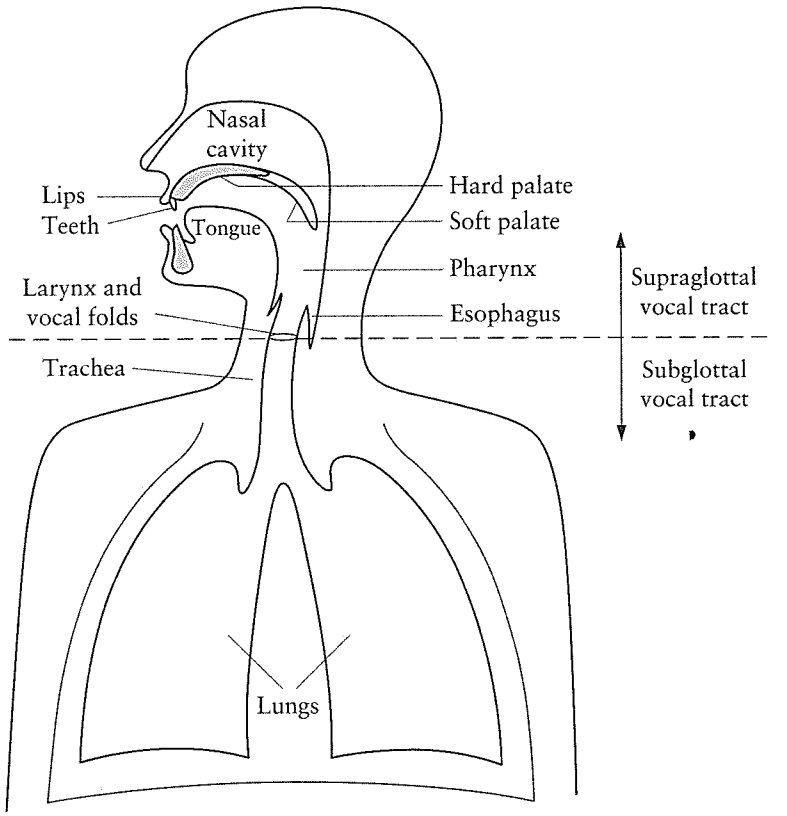

Figure 2.3.1 The organs of speech (greatly simplified and not to scale)

division is based on a functional distinction. The respiratory system below the glottis provides the major energy source for producing speech sounds, while the tract above the glottis determines, in general, the phonetic quality of speech sounds. Most phonetic descriptions of speech sounds are primarily concerned with supraglottal activity.

2.4 Describing speech sounds

Despite the fact that speech is a relatively continuous flow, we are accustomed to thinking of it as sounds, or as sequences of sounds. Conceptually, we treat the flow of articulatory movement as a series of segments. Indeed, this is not just a matter of convenience, for the patterned organization of speech into systematic units and structures is fundamental to its nature, distinguishing speech from mere noise.

The rest of this chapter deals with the ways in which speech sounds can be described, using many of the traditional terms of articulatory phonetics, but showing how these terms often conceal considerable problems of description. The chapter explains the ways in which airflow is generated (2.5) and the role of the larynx as a sound source (2.6) before moving to what are usually thought of as the characteristic and distinctive qualities of vowels and consonants (2.7–2.16).

To a large extent, the segmental nature of speech will remain a convenient assumption in this chapter. Chapter 3 will take up the question of defining and delineating discrete segments and will show that many sounds defy simple segmental assumptions.

2.5 Airstream mechanisms

What Pike (1943) and a number of later writers have called 'airstream mechanisms' provide the sources of energy for generating speech sounds, using airflow and pressure in the vocal tract. Following Pike, we can distinguish three basic mechanisms, namely LUNG AIRFLOW, GLOTTALIC AIRFLOW and VELARIC AIRFLOW.

LUNG AIRFLOW and the respiratory cycle are basic to speech production. In principle, air flowing either into or out of the lungs during the respiratory cycle may be used in generating speech sounds, and the nature of the sound produced will depend on what is happening in the vocal tract above the trachea – on the action of the larynx and on how the rest of the tract is constricted or modified in shape. The two mechanisms (outward and inward lung air) are often referred to as EGRESSIVE PULMONIC and INGRESSIVE PULMONIC. Outward lung airflow is the normal mode: it is easier to control and requires less

overall articulatory effort in sustained speech, largely because speakers can exploit the relaxation pressure available when the lungs are relatively full, and can thus expel air in a slow, controlled fashion.

An egressive pulmonic airstream is the norm in all languages, and languages as diverse as English, Spanish, Indonesian and Chinese use no other mechanism. While it is possible to produce speech using ingressive lung airflow – readers will be aware of the possibility of uttering a gasp or groan or even intelligible vowels with air drawn inward into the lungs – no language in the world seems to use ingressive lung airflow as a distinctive feature of particular speech sounds during normal articulation. There are, however, languages that use glottalic and velaric mechanisms systematically.

The GLOTTALIC AIRFLOW mechanism (sometimes called 'pharyngeal') uses air above the glottis. The glottis is closed, and the larynx is moved up and down the pharynx, under the control of the extrinsic laryngeal muscles, to initiate airflow. Since the glottis is closed, subglottal air is not involved and the larynx thus acts rather like a plunger or piston in a cylinder. If the larynx moves upwards in this way, it can generate an egressive glottalic airstream; and moving downwards, an ingressive glottalic airstream.

Egressive glottalic sounds are commonly known as EJECTIVES, sometimes as 'glottalized stops'. The upward movement of the larynx, with the glottis closed, compresses the air above and forces airflow outward. Readers can attempt such sounds by taking a breath and holding it (thus shutting the glottis), then uttering [p], [t], [k] or [s] without opening the glottis, using only air compressed by raising the larynx. Some speakers of English sometimes produce word-final ejectives, for example at the end of the word *sick*. In the flow of articulation, sounds produced with an egressive glottalic airstream generally precede or follow sounds using normal lung airflow, since the airflow generated by the laryngeal movement is relatively weak and of short duration.

Sounds using an ingressive glottalic airflow are commonly known as IMPLOSIVES. The piston action of the larynx is generally less effective in producing ingressive airflow than egressive, partly because of the difficulty of maintaining a tightly closed glottis during the downward movement of the larynx. As a result, there is often some upward leakage of lung air sufficient to cause involuntary phonation or voicing. According to Ladefoged (1971) this upward leakage may offset the suction action of the downward larynx movement so that there is little or no inward airflow through the mouth, even to such an extent that the net airflow is actually egressive. The sounds can still be counted as implosives, since an important part of their sound quality is due to the effects of rapid larynx lowering during their production.

Ejective stops are found in languages of the Caucasus area, such as Georgian, as well as in a variety of languages of Africa and the Americas. Ejective fricatives are not as common. Implosives are found in a number of African and American languages. The West African language Hausa, for example, has an ejective velar stop (contrasting with pulmonic [k]), an ejective sibilant fricative (contrasting with [s]), and bilabial and alveolar implosives (contrasting with [b] and [d]). Maidu (from central California) has bilabial and alveolar ejectives and implosives (alongside pulmonic [p] and [t]) as well as a velar ejective stop and an

ejective counterpart of the affricate [ts]. Basic discussion of sounds using glottalic airflow can be found in Ladefoged (1971); Greenberg (1966, ch. 2), Maddieson (1984, ch. 7), and Ladefoged and Maddieson (1996, ch. 3) give examples and some observations about the frequency of occurrence of glottalic sounds; and Pinkerton (1986) usefully combines an instrumental analysis of glottalic stops in some languages of Guatemala with a review of Greenberg's predictions about how glottalic sounds function in languages.

VELARIC (or oral) AIRFLOW is generated entirely within the oral cavity, by raising the back of the tongue to make firm contact with the soft palate. Air in front of this tongue closure may then be sealed off by closing the lips or by pressing the sides and tip of the tongue against the roof of the mouth behind the teeth. Although it is possible to generate both egressive and ingressive airflow using this oral air supply, only ingressive airflow is normally used in speech. Sounds produced in this way are commonly known as CLICKS. The simplest form of click is made with the lips, where the action of parting the lips will (with lowering of the jaw) increase oral cavity volume sufficiently to cause a drop in air pressure inside the mouth, causing air to flow in. The action is that of a light kiss. Alternatively, air is trapped in a small chamber created entirely by the tongue itself. The tongue is in effect sucked off the roof of the mouth. When the tongue is moved downwards, the air chamber above it is enlarged and the pressure drop in the trapped air generates a short but quite strong inflow of air as the closure is released. It is this rapid and rather turbulent inflow which causes the characteristic click sound. Click articulation requires complex interaction of the intrinsic and extrinsic tongue muscles, and the tongue can in fact be released in different ways, sufficient to create different click sounds. Readers will be familiar with the kind of click made when the tongue tip is reasonably forward, for the sound is commonly used to express regret or disapproval (usually repeated and sometimes written as *tsk tsk* or *tut tut*); a different click sound, sometimes used by English speakers to urge a horse, is achieved by pulling the tongue down at one or both sides rather than at the tip.

Click sounds are found in rather few languages (about 1 per cent of the world's languages according to Maddieson 1986, p. 115). They are characteristic of the Khoisan languages of the Kalahari area in southern Africa (of which the most famous is probably Hottentot) but are also found in Bantu languages such as Zulu and Xhosa (Westermann and Ward 1933, ch. 19; Ladefoged 1971, ch. 6; Ladefoged and Maddieson 1996, ch. 8). In these languages, clicks are consonants functioning as part of the speech sound system (unlike the *tsk tsk* used to express disapproval, which cannot be considered a speech sound in the same way).

We must also recognize COMBINATORY AIRFLOW PROCESSES, for the muscular systems used in the three airstream mechanisms are autonomous enough to function in partial combination. We noted above, for example, that egressive lung airflow in conjunction with ingressive glottalic airflow results in phonatory action while the larynx is descending. The egressive velaric and egressive pulmonic airstreams can also be activated simultaneously to produce, for instance, click sounds which have a velar nasal sound (as at the end of *sing*) imposed upon them. Such nasal click sounds do occur in languages that exploit the velaric airstream mechanism.

Finally, it should be noted that it is possible to use air from the stomach to generate sound, as in an audible belch. With considerable practice this mechanism, which can be described as egressive esophageal, can be used as a controlled substitute for egressive lung airflow. The technique, sometimes taught to those who have undergone laryngectomy, consists of swallowing air and then belching it out again.

Further discussion of airstream processes can be found in Pike (1943), Catford (1977), Ladefoged and Traill (1980) and Ladefoged and Maddieson (1996).

2.6 Modes of phonation

The term PHONATION refers principally to vocal fold vibration but can also be taken to include all the means by which the larynx functions as a source of sound, not all of which involve vibration of the folds in a strict sense. It is also important to bear in mind that besides this role as a sound source, the larynx has two other functions in speech: it can generate an airstream (yielding glottalic consonants, section 2.5 above) and it can serve as an articulator (in glottal consonants, section 2.10 below).

The complex laryngeal musculature is such that the vocal folds can be manipulated in highly diverse ways, but it is convenient to think in terms of a set of categories known as PHONATION MODES. These categories of laryngeal action are defined not just by observation of the physiology of the larynx, but by reference to distinctions that appear to be relevant in the world's languages. Thus the categories are not simple and direct reflections of different ways of using the larynx, and, as in many other areas of phonetic description, not all the details of physiology are relevant to the categories that are appropriate for describing speech.

Catford (1964, 1968, 1977) is responsible for a highly detailed set of categories: his emphasis is on what he terms 'anthropophonic' possibilities, that is, on comprehensive coverage of all the articulatory possibilities. Laver (1968, 1980) offers a complete theoretical and practical descriptive system for laryngeal (and other) aspects of voice quality. Both accounts exploit combinations of a series of basic laryngeal settings. Catford, for example, defines some 13 phonation modes derived from four types of glottal stricture and three locations of phonatory activity. Other linguists (such as Halle and Stevens 1971, Ladefoged 1971 and Ladefoged and Maddieson 1996) work with rather fewer categories, as it is evident that real languages actually do not exploit all of the distinctions which a phonetician may recognize on articulatory or physiological grounds. The following account focuses on the distinctions that do seem relevant in language, and recognizes five phonation modes, namely VOICELESSNESS, WHISPER, BREATHY VOICE, VOICE and CREAK. The distinction between voiceless and voiced sounds applies in a high proportion of the world's languages (though it is

certainly not universal); distinctive use of breathy voice and creak is much less common; and whisper could arguably be omitted as nonlinguistic, but it is included here both to underline its difference from voicelessness and breathy voice, and because of its widespread use (as in English) as a distinctive style of speech rather than as a feature of specific sounds.

VOICELESS means the absence of any phonation. The vocal folds are held far enough apart to allow a laminar (or nonturbulent) airflow through the glottis. If the airflow is more than moderate, even this open setting of the glottis will generate turbulence (which in fact allows the glottis to function as a sound source for a glottal fricative such as the [h] in English *hand* or *head*). Catford's figures (1977) suggest that voiceless articulation is maintained provided that airflow does not exceed 200–350 cm^3 per second (depending on the degree of glottal opening). Vocal fold abduction is largely a function of the posterior cricoarytenoid muscle action, and the opening of the glottis is usually greater in the voiceless mode than in any other mode used in speech. Ladefoged (1971) suggests that the opening for voiceless articulation is similar to that required in normal breathing. Voiceless sounds in English include the stops [p] (as in *pea*), [t] (as in *tea*) and [k] (as in *key*), and fricatives [f] (as in *fee*), [θ] (as in *theme*) and [s] (as in *see*). Many of the world's languages have similar sounds contrasting with their voiced counterparts: the distinction between voiceless [f] and [s] and voiced [v] and [z], for instance, is found in languages as diverse as French, Greek, Russian, Hungarian, Turkish, Vietnamese and Zulu.

WHISPER requires far greater constriction than the voiceless setting of the glottis, and it is generally achieved by adducting the ligamental vocal folds while maintaining an opening between the arytenoid cartilages, through which the bulk of the airflow is forced. This setting can be created by the lateral cricoarytenoid muscles (contributing to medial compression of the ligamental folds) and the posterior cricoarytenoid muscle (contributing to abduction of the arytenoids). Adduction of the false vocal folds may also help to narrow the glottal airflow path, and to inhibit true vocal fold vibration (Sawashima et al. 1969).

The characteristic consequence of the whisper setting is that there is significant turbulence at the glottis. This functions as a sound source which can then be modified by articulatory activity in the supraglottal vocal tract. As the area of glottal opening is small, this mode can provide turbulence with relatively low airflow rates (from about 25 cm^3 per second according to Catford 1977). Whisper thus exploits a usable sound source without demanding a large air supply from the respiratory system; but it does also require considerable overall laryngeal tension. Readers should be able to verify the degree of tension by changing back and forth between whisper and quiet breathing.

In BREATHY VOICE, normal vocal fold vibration is accompanied by some continuous turbulent airflow. This occurs when glottal closure during the vibratory cycle is not complete (hence the term 'breathy'). Usually the arytenoid cartilages remain slightly apart while the ligamental folds vibrate; in some speakers, ligamental fold closure may also be weak or incomplete, accounting for part or even most of the turbulent air leakage.

There is some terminological inconsistency around this kind of phonation. We retain the term 'breathy voice', which is relatively widespread and has

reasonably obvious relevance; but Heffner (1964) and Ladefoged (1971, 2006) use the term 'murmur', and Catford (1968, 1977) and Laver (1980, 1994) use 'whispery voice'. For Catford, 'breathy voice' is a phonation mode with a very high rate of airflow, in which, according to his description, the vocal folds 'flap in the breeze'. See Sprigg (1978) for a general review of phonation description (including some criticism of Catford). See also Gordon and Ladefoged (2001) for a review of relevant physical measures of different phonation types.

Several languages of South Asia make a systematic distinction between breathy voice and normal voiced phonation: in transliterations of Hindi and Urdu, for example, spellings such as *bh* and *gh* indicate plosives with breathy voiced release which are distinct from voiced *b* and *g*. In some languages, such as Tamang (a Sino-Tibetan language spoken in Nepal), vowels may have distinctive breathy voice. In English, breathy voicing is not exploited in the same way but is an identifiable feature of some speakers, either as part of their personal voice quality or as a result of some laryngeal disorder.

VOICE refers to normal vocal fold vibration occurring along most or all of the length of the glottis. Physiologically, there is a continuum of subtypes within this category (Ladefoged 1971). At one end of the continuum, approaching breathy voice, the muscles controlling vocal fold adduction are relatively relaxed; at the other end, tension in the musculature begins to limit the vibration of the folds and voice verges on laryngealized or creaky voice (described below). In a language such as English, individuals normally exploit a range of laryngeal muscle settings, constrained by such factors as the degree of vocal effort needed (e.g. shouting versus very quiet speech) and physiological and emotional state (e.g. tiredness or excitement). The consequent variation in voice quality can be described impressionistically as ranging from 'dark' or 'mellow' (the most relaxed end of the range of muscle settings) to 'bright' or 'sharp' or 'hard' (the most tense end of the muscle setting range).

All languages have voiced sounds, and voicing can be considered normal for sounds such as vowels and nasal and lateral consonants. In English, for example, vowels are always voiced, and nasal and lateral consonants are voiced unless devoiced by assimilation (as in e.g. *play* or *clay*, where the [l] may be voiceless by assimilation to the preceding voiceless stop). But the precise settings of the larynx that can be regarded as producing 'normal voice' depend not only on the language, or regional or social dialect, but also on the individual (Laver 1968, Laver and Trudgill 1979).

CREAK is a phonation mode characterized by low frequency vibration of the vocal folds. The folds open only for a very short time and often quite irregularly from cycle to cycle of vibration. It has also been variously described as 'laryngealization' (Ladefoged 1971, 2006, p. 143), 'pulsation' (Peterson and Shoup 1966a), 'vocal fry' (Wendahl et al. 1963) and 'trillization' (Pike 1943, Sprigg 1978). There is some uncertainty among researchers about exactly how creak is produced, but the majority view is that the arytenoids are tensely adducted, and that only the anterior part of the ligamental folds vibrates. According to Catford (1977), subglottal pressure and airflow rates may be quite low, and the ligamental folds tightly closed but not greatly tensed.

In addition to these five phonation modes we must allow for COMBINATORY PHONATION MODES. These include BREATHY CREAK, in which creak is accompanied by some turbulent air leakage to produce breathiness, and VOICED CREAK, in which creak and normal voice are combined. Voiced creak is sometimes referred to as 'laryngealization', but this term should be treated with caution, as some writers use it to describe simple creak, and others use it to refer to a complex articulation in which complete glottal closure follows or accompanies some other articulatory gesture. For full details of the anatomy of the larynx and its various phonatory settings, see sections 6.5 and 6.6 below.

2.7 Vocalic sounds

What we commonly think of as vowel sounds are better described, when considering their articulation, as vocalic sounds. (It is often convenient to use the word 'vowel', but for some purposes it is necessary to distinguish between vowels and vocalic sounds, and we shall come to the reasons for that in chapter 3.) Vocalic sounds are produced by egressive pulmonic airflow through vibrating or constricted vocal folds in the larynx and through the vocal tract, and the sound generated at the larynx is modified by the cavities of the tract. The size and shape of the tract can be varied, principally by positioning of the tongue and lips; and as the tract is varied, so the perceived phonetic quality of the vocalic sound is altered. Thus the two most fundamental articulatory manoeuvres in producing various vocalic sounds are the shape and position of the tongue, and the shape and degree of protrusion of the lips. It is the tongue that largely determines the geometry of the oral and pharyngeal cavities, and the lips that control the shape and area of the front of the vocal tract. Lip protrusion also provides a means of extending the overall length of the vocal tract.

The major challenge in describing the articulation of vocalic sounds is to define the position of the tongue. The tongue moves within a spatial continuum without making any significant constriction in the area surrounding the midline of the oral cavity. As a result, we cannot locate a specific point of constriction or blockage, and phoneticians have had to struggle to devise a satisfactory way of plotting the position of the tongue (Ladefoged 1967, ch. 2).

Traditionally, vowels are plotted on a two-dimensional diagram representing the articulatory space: the vertical axis is tongue HEIGHT, and the horizontal axis is tongue FRONTING (or backness or retraction). There is no handy landmark on the tongue to serve as a point of reference in this mapping, but the traditional procedure has been to try to locate the highest point on the dorsum of the tongue. The height and fronting of this point are then plotted relative to some external reference point such as the atlas vertebra. An early example of the procedure is found in the frontispiece photographs in Jones (1960, first published 1918). The articulatory positions of some Australian English vowels, defined similarly, are shown in figure 2.7.1 (based on lateral X-ray photographs by Bernard 1970b). Both Bernard and Lindau (1978) describe this measurement

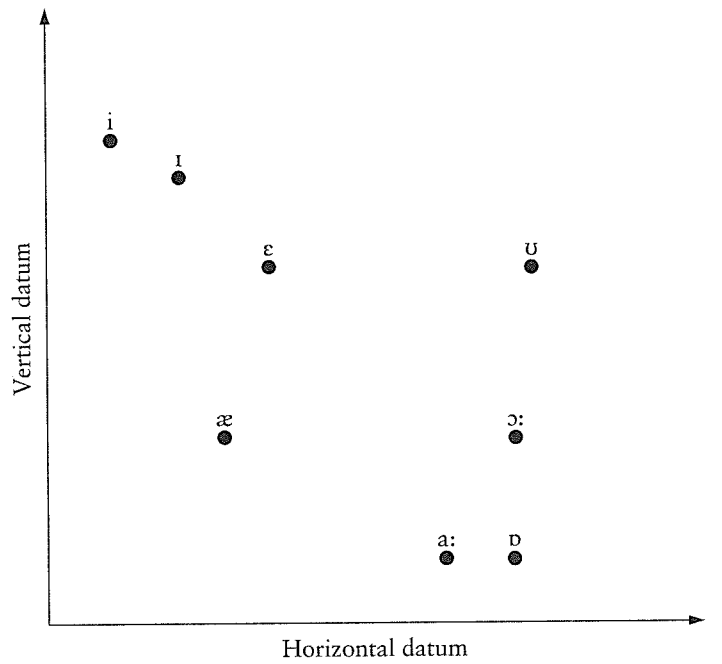

Figure 2.7.1 Articulatory positions of some Australian English vowels
Adapted from: Bernard 1970b.

procedure in detail; and Lindau extends it to account for other aspects of tongue posture.

Assuming that this method gives a valid measure of vocalic articulation, it is, however, impractical to take X-ray pictures for every vowel in every language and dialect that we would like to describe. As an alternative, we can use some form of acoustic analysis of the sound itself (chapter 7 below): this has sometimes been impractical because of the equipment needed for recording and analysis, but modern computer technology and the ready availability of signal processing software are making this kind of analysis much more feasible (see section 7.14 below). Another possibility is to base the description on auditory impressions. The disadvantage here is precisely that it is to some degree impressionistic: the observer needs to be well trained in phonetics, and even then will still be influenced by conventional terminology and by linguistic experience, since none of us is ever entirely free of perceptual bias shaped by the language(s) which we happen to speak.

In an effort to bring accuracy and objectivity into impressionistic vowel descriptions, nineteenth-century phoneticians such as Alexander Melville Bell tried to define standard categories of vowel quality and associated articulatory positions. The most successful outcome of this idea, and one still in use for vowel description, is the set of CARDINAL VOWELS devised by Daniel Jones. These vowels are intended to serve as standard reference points, or 'cardinal' points in Jones's terminology.

The cardinal vowels are not drawn from any particular language or languages but are derived from a kind of grid imposed upon the space in which the tongue moves. There are 16 cardinal vowels in all, eight primary and eight secondary. In each set of eight there are two vowels which represent the outer limits of vocalic articulation, the boundaries beyond which vocalic sounds cannot be produced: if the tongue exceeds these boundaries, it will create constriction in the vocal tract sufficient to generate a consonant rather than a vowel. In the primary cardinal vowel set, the first reference vowel is cardinal 1, produced with the tongue as high and as far forward in the mouth (towards the hard palate) as it is possible to go without causing audible friction. (The nearest example in English is an extremely raised and fronted form of the vowel in *heed*.) The second reference vowel is cardinal 5, produced with the tongue as low and retracted as possible. (The nearest English example is an extremely lowered and retracted form of the vowel in *hard*.) The reason for choosing these two vowels as starting points is that they are the easiest (or perhaps least difficult) to locate by the feel of the tongue. From cardinal 1, Jones then defines cardinals 2, 3 and 4 as vowels for which the tongue is still fronted but is lowered in equal steps. Thus 1 and 2, 2 and 3, and 3 and 4 are supposed to be auditorily equidistant. The back vowels of the series are similarly formed, starting from cardinal 5 and raising the tongue in a retracted position such that 5, 6, 7 and 8 are again equally spaced from lowest to highest.

Cardinal vowels 1 to 5 are produced with the lips in a neutral or spread position (most spread for 1 and progressing to neutral for 4 and 5). Cardinals 6 to 8 are produced with the lips rounded. The eight secondary cardinal vowels are produced exactly as the primary set, except that the lip positions are reversed: cardinal 9, for example, has the same tongue position as cardinal 1, but with lips rounded instead of spread; cardinal 16 has the same tongue position as cardinal 8, but with lips spread instead of rounded.

The cardinal vowels are thus intended to represent the most peripheral tongue positions for vocalic sounds. They stand, so to speak, on the boundary of vocalic articulation, and it should be possible to locate any vowel in any language somewhere within the area encompassed by this boundary. Jones took the tongue positions for cardinals 1, 4, 5 and 8 from lateral X-ray photographs of his own productions of these vowels. He then constructed a quadrilateral with these four vowels at the corners (Jones 1960, pp. 36–7). The vowel quadrilateral is irregular – somewhat like a diamond tilting to the left – but a slightly simplified version of it (figure 2.7.2) is now standard.

The vowels of particular languages are commonly placed on a vowel quadrilateral to locate their phonetic qualities relative to the cardinal vowels, but this strategy of description must be treated with caution. The fundamental worry about the cardinal vowel system is that it confuses articulatory and auditory properties. Note that the two reference points in the system (cardinals 1 and 5) are established on physiological grounds – they are at the outer limits of tongue movement for vocalic articulation. On the other hand, intermediate vowels are determined by what Jones calls equal 'acoustic' (i.e. auditory) intervals along the continuum. Despite this, Jones implies that the tongue positions of the cardinal vowels also progress in equal steps, and he describes the cardinal

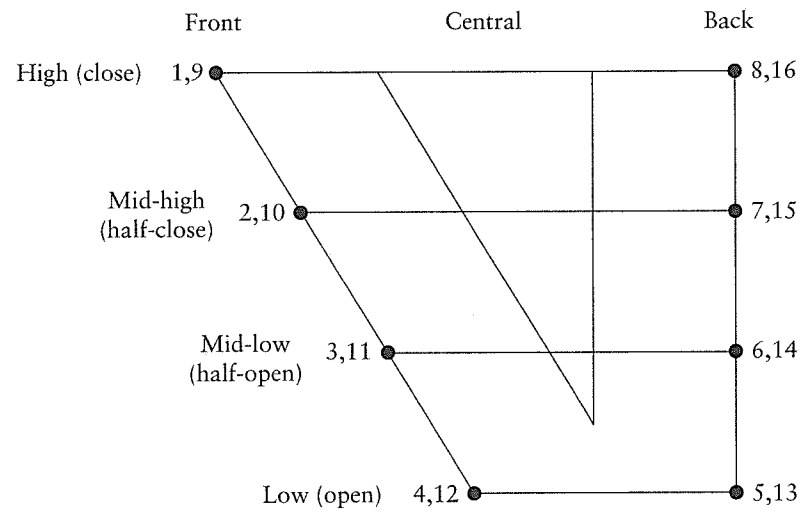

Figure 2.7.2 The cardinal vowel diagram

vowel diagram itself in terms of tongue position (as do many linguists after him). Now it may seem reasonable to suppose that changes in articulatory setting and changes in auditory quality go hand in hand; and that one can therefore judge articulatory position from auditory perception. In fact Ladefoged (1967) has shown that the assumption is not fully warranted. He examined X-ray photographs of a complete set of cardinal vowels (published in 1929, relatively soon after Jones's original work) and measured the tongue positions. His measurements reveal that the front vowels (cardinals 1–4) are indeed roughly equidistant, but not the back vowels (5–8): tongue height is actually identical for cardinals 6 and 7, which are also much farther from 8 than they are from 5. Lindau (1978) provides data to show that back vowels in natural languages similarly fail to conform to the cardinal idealization.

A second difficulty with the cardinal vowel system is that the specifications of tongue position suggest an invariant tongue position for each vowel quality. But, as Lindau (1978) has pointed out, X-rays of vowels in actual languages show that speakers generally have several possible ways of producing a given auditory vowel quality. Moreover, this is not just a matter of variation in tongue posture, for vowel quality is also affected by changes in jaw aperture and larynx height. Experimental investigations by Lindblom and Sundberg (1971), Ladefoged et al. (1972), Riordan (1977), Lindblom et al. (1979) and Johnson et al. (1993) all clearly show that speakers are capable of a considerable degree of compensatory articulation to produce a single desired auditory result in vowel quality. There is thus no reason to assume a one-to-one matching of articulatory position and auditory quality.

A third problem concerns the definition of tongue position. In the classic formulation, tongue height is taken to mean the height of the point which is closest to the roof of the mouth. But tongue position could be measured in various ways, and there is no principled reason why the location of maximum

tongue height should correspond directly and systematically to vowel quality (Lindau 1978, Wood 1979). Recent research suggests that it is the location of the major constriction formed by the tongue, rather than tongue height itself, which is a much more direct determinant of perceived vowel quality. Overall, it appears that the measures needed in vowel descriptions are rather more complex than the traditional one of tongue position; and this helps to explain some of the weaknesses in the supposedly physiological basis of the cardinal vowel system (Ladefoged 1967, Harshman et al. 1977).

Given these difficulties, the cardinal vowels are best taken to be auditory qualities rather than articulatory specifications. Understood in that way, they can serve a useful purpose in helping phoneticians to identify vowel qualities and in bringing some measure of objectivity into auditory judgements. The continuing use of articulatory labels for auditory qualities is unfortunate, but there is no easy alternative, since we lack a well-developed perceptual terminology. The fact that many phoneticians have used the system with a considerable degree of consistency is largely due to thorough training. Jones himself stressed 'ear training' and the importance of learning the cardinal vowels from a competent teacher, or at second best from a recording. The only 'standard' recording of the cardinal vowels is by Jones himself, and he trained a number of students at University College London, many of whom later became senior phoneticians in other British universities, so that something of a direct oral tradition has been maintained.

The lip position of vocalic sounds raises far fewer difficulties than tongue location, if only because the lips are externally visible. We have already seen that cardinal vowels may have SPREAD, NEUTRAL or ROUNDED lips, and figure 2.7.3 illustrates these three settings. The difference between spread and neutral lip positions can generally be associated with vowel height: while a high vowel may have spread lips (as cardinal 1 does), a lower vowel will tend to have a more neutral lip posture, chiefly because the larger jaw aperture will tend to produce a more neutral lip position (unless the lips are deliberately rounded). For this reason, and because few if any languages actually exploit a distinctive difference between spread and neutral lips, the two positions are often united under the label UNROUNDED, which underlines the contrast with the ROUNDED lip position. Lip rounding may include some degree of lip protrusion, and there is commonly more protrusion in back rounded vowels than in front rounded vowels. According to Catford (1977), this may be motivated by the need to preserve the auditory impression of fronting in front rounded vowels.

Conventional symbols for the primary and secondary cardinal vowels are listed in table 2.7.1. It should be emphasized again that the cardinal vowels are not derived from English or any other language: the sample words are intended only as helpful approximations. Figure 2.7.4 shows the vowel symbols of table 2.7.1 on a cardinal vowel diagram. Figure 2.7.5 shows some English vowels as phonetic symbols mapped on to a cardinal vowel diagram. The vowels are based on Gimson (1980) and represent British RP.

Some additional modifiers, or DIACRITICS, serve two functions: the first is to locate a vowel within the auditory space, relative to the cardinal vowel closest to it; the second is to indicate the relative length of the vowel. Some of the commonly used diacritics are as follows (where V represents any vowel symbol):

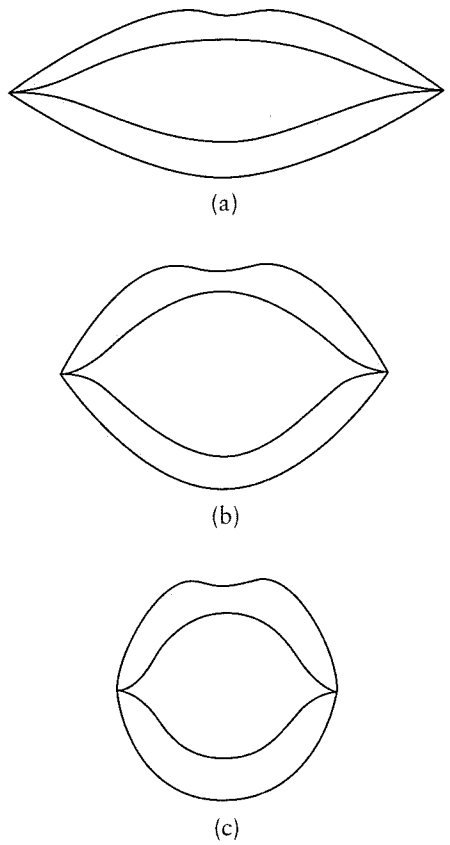

Figure 2.7.3 Lip positions in vowel articulation: (a) spread; (b) neutral; (c) rounded

V̝ – raised with respect to V
V̞ – lowered with respect to V
V̟ – fronted or advanced with respect to V
V̠ – retracted with respect to V
V̽ or V̈ – centralized with respect to V
V. – half long, or slightly lengthened
V: – long

The use of symbols with diacritics to represent fairly precise estimations of vowel quality is sometimes known as NARROW phonetic transcription. It is not always necessary or possible to include all the detail of a narrow transcription, and it may be sufficient to make a BROAD transcription using the nearest appropriate cardinal symbols with few or no diacritics. For this reason many of the symbols have come to have conventional values in particular languages: cardinal 14, for example, is regularly used to represent the English vowel of *but* and *luck*, even though this vowel is central rather than back in many varieties of English, including RP and Australian English. Cardinal 8 may likewise be used to transcribe the high back vowel of Japanese (which is actually unrounded

Table 2.7.1 Cardinal vowel symbols

Cardinal vowel no.	*Symbol*	*Lip position*	*Sample words illustrating approximate vowel quality*
1	[i]	unrounded	English *beat*, French *si*
2	[e]	unrounded	French *chez*, Italian *che*
3	[ɛ]	unrounded	English *bet*, German *wenn*
4	[a]	unrounded	English *spa*, French *la*
5	[ɑ]	unrounded	Dutch *dam*, French *las*
6	[ɔ]	rounded	English *hawk*, French *côte*
7	[o]	rounded	French *beau*, Italian *lo*
8	[u]	rounded	French *ou*, German *gut*
9	[y]	rounded	French *tu*, German *für*
10	[ø]	rounded	French *eux*, German *Goethe*
11	[œ]	rounded	French *heure*, German *Götter*
12	[Œ]	rounded	(not distinctive)
13	[ɒ]	rounded	English *hock*, Dutch *dom*
14	[ʌ]	unrounded	English *but*, *luck*
15	[ɤ]	unrounded	Vietnamese *ô*
16	[ɯ]	unrounded	Japanese *u*, Vietnamese *ư*

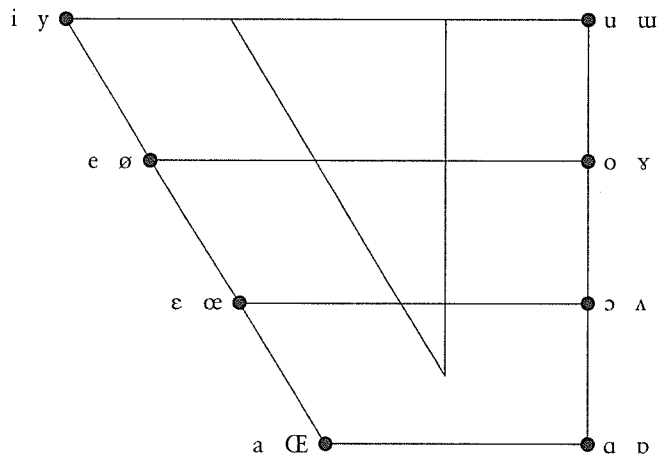

Figure 2.7.4 Cardinal vowel symbols located on the diagram of figure 2.7.2

rather than rounded) and the vowel of Australian English *boot* and *food* (which is rather more central than back).

The 16 cardinal vowel symbols have been supplemented by some additional symbols (table 2.7.2). These are technically redundant, since they could be replaced by cardinal vowels with diacritics; but they represent particular vowels for which it is judged convenient to have a distinct symbol. Contrary to the spirit of the cardinal system, some of them are conventionally understood to be inherently short or long.

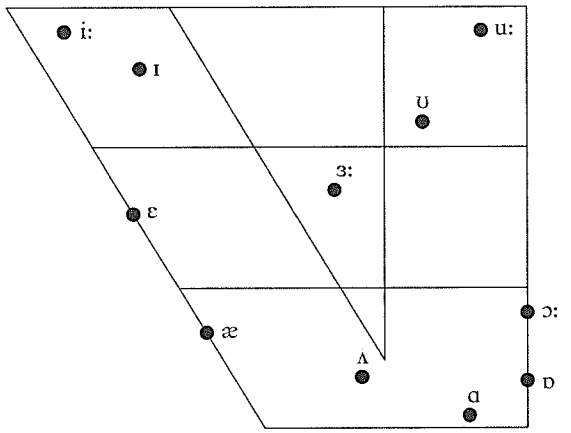

Figure 2.7.5 English vowels: typical RP values

Table 2.7.2 Additional vowel symbols

ɨ	high central unrounded vowel as in New Zealand English pronunciation of *pit* or *six*; value of Russian ЫІ
ʉ	high central rounded vowel as in Scottish pronunciation of *put*
ɪ	centralized version of cardinal 1, usually understood to be a short 'lax' vowel, as in RP English *pit*
ɩ	alternative symbol for 1
ʊ	centralized version of cardinal 8, usually understood to be a short 'lax' vowel, as in RP English *put*
ɷ	alternative symbol for ʊ
ə	central unrounded vowel, known as schwa: used in RP and similar varieties of English to represent the unstressed or 'indeterminate' vowel, as initial in *about* or final in *China*; also used to represent the (stressed) vowel of *cup* or *luck* as pronounced in North American English
ɜ	long central unrounded vowel, equivalent to lengthened schwa, as heard in RP English *bird* or *hurt*

In the description of languages, it is sometimes sufficient to represent the vowel sounds in a general auditory space without following the precise format of the cardinal system. Such displays may retain the quasi-articulatory dimensions of height and fronting, but are often intended to show relative differences in phonetic quality among members of a vowel *system* in a particular language or dialect. For this purpose, much of the phonetic detail can be judged irrelevant. Figure 2.7.6 displays the vowel system of Australian English in such a way.

Vowel systems vary greatly in their complexity from language to language. English happens to be relatively rich in vowel contrasts, with the added complexity that the vowel system is by no means uniform across the English-speaking world. Australian English, as shown in figure 2.7.6, represents one of the richer systems; note for instance that the distinction between the vowels of

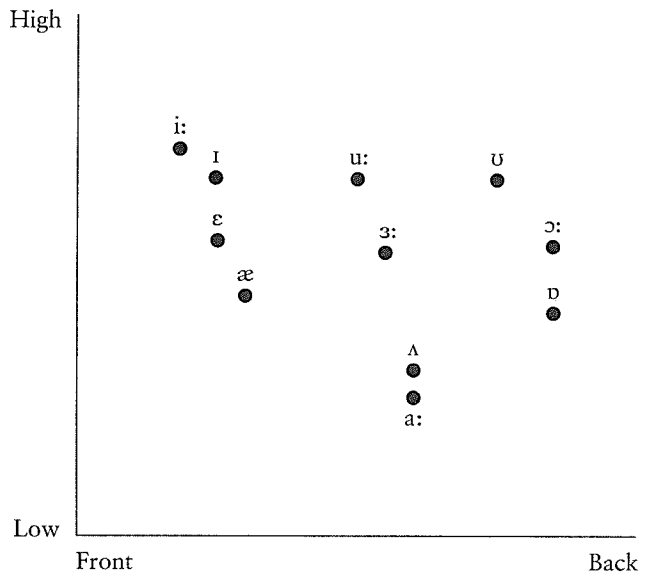

Figure 2.7.6 Australian English vowel system

look and *Luke* is not universal, notably not in Scotland. RP (and the English of south-eastern England in general) is systematically comparable to Australian, although the precise quality of many of the vowels is quite different.

Most of the world's languages have rather fewer vowels, and some, including Classical Arabic and some Australian Aboriginal languages, have only three distinctive vowels. In a three-vowel system, the vowels are usually towards the outer edges of the vowel space, in the general region of cardinals 1, 4/5 and 8, i.e.

i u
a

Five-vowel systems are widespread and are often similarly distributed. A common pattern (found, for example, in Spanish, Modern Greek, Maori and other Polynesian languages, and Swahili and some of the other Bantu languages of eastern and southern Africa) can be represented as follows:

i u
e o
a

Patterning is also revealed in other ways. For example, vowel length is often exploited in such a way that each short vowel is matched by a long vowel. In several Australian Aboriginal languages, for instance, we have

i i: u u:
a a:

So far as lip rounding is concerned, languages appear to favour unrounded lip position for front vowels and rounded for back vowels. (This serves to enhance

auditory difference between front and back vowels, although the rounding of back vowels is not always very prominent.) Few languages distinguish unrounded back vowels from rounded back vowels, and where rounded front vowels occur, they are normally found in addition to front unrounded vowels and not instead of them. German, for example, has the following long vowels (and some other Western European languages such as French and Dutch are broadly comparable in that they distinguish front unrounded, front rounded and back rounded vowels):

i: y: u:
e: ø: o:
a:

Lindblom (1986) provides a brief but useful survey of 'some facts' about vowel systems as well as some discussion of how languages exploit the 'vowel space'. His paper includes references to both classic and recent work on universal aspects vowel systems. Lindblom suggests, however, that vowels are not necessarily evenly dispersed over the available 'space', but may follow the principle of 'sufficient contrast'. In other words, the vowels of languages with small vowel inventories tend to occupy a more compact space than those of languages with large vowel inventories.

In most languages vowels are normally voiced. Conventions for symbolizing other modes of phonation (section 2.6 above) are not well established, but breathy voice may be signalled by two dots beneath the main symbol – e.g. [a̤] – and creak by a tilde beneath the main symbol – e.g. [a̰] (cf. Ladefoged 2006, p. 152). A voiceless whispered vowel may be symbolized by a diacritic used to indicate voicelessness more generally, namely a small subscript circle – e.g. [ḁ].

Vocalic sounds are normally produced with an oral airstream. That is, the velum is raised, preventing major airflow through the nasal cavities, although there may be some nasal 'leakage' if relatively little muscular effort is used to raise the velum. By contrast, a vowel may be distinctively NASALIZED when the velum is deliberately lowered to ensure substantial airflow through the nasal cavities. The nasal cavities are then said to be coupled to the oral and pharyngeal cavities, and the effect of the coupling is an audible nasalized quality. Nasalized vowels are found in a fair number of languages (French, Portuguese, Hindi and Burmese among others), usually as a subset of the oral vowels. French, for example, has four nasalized vowels alongside some 12 oral vowels: the four nasalized vowels are heard in the phrase *un bon vin blanc* 'a good white wine', and the contrast between oral and nasalized vowels is evident in pairs of words such as *beau* 'fine' versus *bon* 'good', and *bas* 'low' versus *banc* 'bench'. (For further remarks on nasalization see section 3.3 below.)

There are other articulatory variables which affect vowel quality. One commonly cited is tenseness, although the notion that vowels can be validly described as tense or lax is controversial. Tenseness is generally described as an overall tightening of vocal tract musculature, associated with definite or forceful articulatory action. A tense vowel is therefore likely to be longer and

more peripheral in quality than a corresponding lax vowel. Examples often quoted from English (especially American English) are the tense vowels in *beat* and *boot* compared with their lax counterparts in *bit* and *put*. Stevens et al. (1966) report instrumental evidence to support the nature of the distinction, and MacNeilage and Sholes (1964) note greater tongue muscle activity in tense vowels. Appealing to cine-radiographic evidence, Perkell (1969) suggests that the tongue attains a more stable and definite position in vowels that are judged to be tense. He also comments, however, that it is not clear that there is a distinct articulatory mechanism or strategy to account for what is impressionistically reckoned as tenseness. Ladefoged implies that tenseness may be a matter of pharynx width: if the tongue root is moved forward, it is possible to widen the pharynx without any effective alteration in tongue height. When the pharynx is widened in this way, the tongue is bunched along its length and therefore – on one interpretation of the term – 'tensed'. Ladefoged draws on data from Akan which, like a number of other West African languages, has two sets of vowels apparently distinguished by pharynx width (Lindau 1979; Ladefoged and Maddieson 1996, pp. 300–6; Ladefoged 2006, pp. 223–4). But he also notes that Akan speakers seem to use different methods of widening the pharynx: some advance the tongue root, others rely more on lowering the larynx.

We will avoid the simple labels 'tense' and 'lax' while noting that something like WIDENED PHARYNX or ADVANCED TONGUE ROOT is essential in the description of at least some of the world's languages. The labels 'tense' and 'lax' should be treated cautiously, given their apparent articulatory implications, for vowels that are often described as tense and lax may be distinct in several ways: the English vowels in *beat* and *bit* (in some varieties of English) may differ in pharynx width and perhaps in tongue tension, but they also differ in length and tongue position. It may well be appropriate, in the description of a specific language, to subsume a number of differences under the tense–lax distinction. But in that case, 'tense' is likely to mean different things in different languages (or may even mean different things for different vowels within one language), and it becomes all the more unreliable as an articulatory label. Ladefoged and Maddieson (1996, ch. 9) present a summary of the different phonetic parameters and articulatory settings that are found across vowel systems.

2.8 Duration and glide in vocalic articulations

We have already referred briefly to vowel LENGTH (or duration) in the preceding section. To some extent, length is dependent on, or conditioned by, other factors, in particular by the quality of the vowel and by consonants adjacent to the vowel.

All other things being equal, certain vowels tend to be longer than others. Lehiste (1976) speaks of the INTRINSIC duration of a vowel. Thus low vowels

tend to be intrinsically longer than high vowels, because of the greater overall articulatory movement and biomechanical effort required to produce the lower vowels, particularly where major tongue and jaw movements are needed.

The effects of adjacent consonants on vowel duration are rather more complex, and it is not always easy to distinguish the influence of an adjacent consonant from a feature of pronunciation that is simply peculiar to the language concerned. In English, for example, vowels followed by voiced stops and fricatives are considerably longer than those followed by voiceless consonants: compare *feed* and *feet* or *fad* and *fat*. But while this may strike speakers of English as a natural and inevitable effect, lengthening before voiced consonants turns out not to be a universal feature – at least not to the same extent as in English. On the other hand, the point of articulation of neighbouring consonants does seem to have an inevitable effect on the duration of a vowel. If a consonant involves tongue movement, more time will be needed to establish the consonantal articulation, and the adjacent vowel will be longer. Thus vowels are likely to be longer before alveolars or velars than before bilabials, for example.

Length is not merely a conditioned feature of vowels, however, but can also function distinctively. Sometimes it works alongside other features. Thus in English – or at least some varieties of English – length is one of the factors differentiating *heed* from *hid* and *wooed* from *wood*. Sometimes length is the crucial distinguishing feature. Bernard (1967) has shown that the distinction between the long vowel of *calm* and *heart* and the short vowel of *come* and *hut* in Australian English is entirely a matter of duration. In some languages length is exploited rather more systematically than this. In languages such as Finnish and Hungarian, for example, there are two matching sets of long and short vowels: every short vowel has a long counterpart and every long vowel a short counterpart (although vowel quality may not be exactly identical across each pair of vowels).

Where vowel length is distinctive in this way, it is relative duration that matters rather than absolute duration. The length of any vowel will be in some measure dependent on its quality and context, and there is no minimum length for a long vowel or maximum length for a short vowel. If two vowels contrast with each other in length, what matters most is their duration relative to each other in comparable contexts. Thus the English short vowel in *hid* and *bid* is longer than in *hit* and *bit* (because of the effect of the voiced [d]) but it is still short relative to the long vowel of *heed* and *bead*; while the long vowel of *heed* and *bead* is shorter in *heat* and *beat* but still long relative to *hit* and *bit*. Bernard's studies (1967, 1970a) demonstrate this point for Australian English. Measurement of vowel duration thus reveals various degrees of length intermediate between the shortest and longest values. In general, however, the functional relativity of length is such that it is rarely if ever necessary to recognize more than two values in any particular language: functionally, vowels are either short or long (or neither if length is not distinctive in the language). Nevertheless, this simple conclusion about vowels must be set in a wider context, for syllabic organization and prosody also affect the way in which duration is exploited – in English, for instance, stressed syllables are normally longer than unstressed.

Simple vocalic sounds have a steady state articulation; that is, the tongue, lips and jaw are meant to achieve – however briefly – a stable configuration, commonly called the TARGET configuration. If produced in isolation, as in a demonstration of cardinal vowels in a phonetics class or in a singing exercise, a vowel can be prolonged without any appreciable change in quality. In normal connected speech, however, there is almost always some articulatory movement at the start and end of a vocalic sound. At the beginning of a vowel, the tongue and lips may be moving away from the configuration of the preceding consonant, and at the end, they may similarly be anticipating the gestures needed for a following consonant. For reasons such as these, the vowel target is normally preceded and followed by rapid TRANSITIONS. These transitions actually play a significant role, as they seem to be important cues in our perception of speech, but they do not disturb our impression that certain vocalic sounds have a single stable auditory quality. A vowel which meets this condition can be termed a PURE VOWEL.

It is also possible to make a deliberate movement of the articulators, particularly the tongue, during a vowel. Here the movement is not a direct response to the articulatory demands of adjacent consonants, but is usually somewhat slower, and constrained within the articulatory repertoire of the language concerned. The resulting change in auditory quality, either before or after the main vowel target, is known as an ONGLIDE or OFFGLIDE. The occurrence of such glides is quite language-specific, but the articulatory movement involved is often towards or from a generally centralized position. The range and direction of a glide, relative to the target, can be conveniently displayed on a cardinal vowel diagram. In transcription, an onglide can be represented as a superscript before the vowel, an offglide as a superscript after the vowel. For example, the vowel in *fee* has a noticeable onglide in many varieties of English and can be transcribed as [fᵊi]. The vowel in *four* may have an offglide (for example in conservative RP or in the southern USA) and can be symbolized as [fɔᵊ]. Figure 2.8.1 shows the two glides on a cardinal vowel diagram.

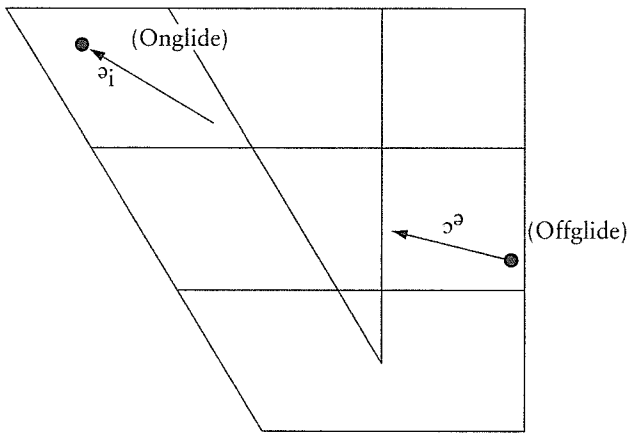

Figure 2.8.1 Vowel onglide and offglide

In some vocalic sounds, the glide component is so prominent that the vowel no longer has a single identifying vowel target value, even though it is still heard as a single sound. Such sounds are DIPHTHONGS. Articulatory movement, particularly of the tongue, occupies a substantial portion of a diphthong, which can be defined in terms of two vocalic targets that determine the range and direction of the glide between them. Diphthongs may be mapped on a cardinal vowel diagram, and are transcribed by a digraph consisting of the two vowel symbols which best represent the two targets.

The vowels in *high* and *hear* are diphthongs in RP, and can be symbolized as [aɪ] and [ɪə]. (Many varieties of English have a comparable diphthong in *high* but not in *hear*: many Scottish and American speakers, for instance, will pronounce *hear* with a pure vowel followed by a consonantal *r*.) Traditionally in English phonetics, diphthongs such as [aɪ] produced with a tongue movement from a mid or low to a high position are known as CLOSING DIPHTHONGS (i.e. moving to a closer tongue position), while those like [ɪə], produced with a tongue movement from a peripheral to a central position, are known as CENTRING DIPHTHONGS. Figure 2.8.2 shows these two examples on a cardinal vowel diagram. Diphthongs vary widely in their total duration, and, like pure vowels, are influenced by their environment. Functionally, they count as long vowels, and just as the long vowel of *heat* is even longer before a voiced stop in *heed*, so the diphthong of *height* is longer in *hide*. The measured duration of English diphthongs ranges from about 150 to 400 ms.

There is no simple way of deciding the difference between a pure vowel with onglide or offglide, and a diphthong. The two targets of a diphthong can have very unequal durations, and the duration of the glide relative to the total length of the diphthong is also variable. This means that two diphthongs can have similar targets and comparable total duration but vary in their auditory quality. One consequence of this is that the digraph notation is only approximate, although for greater accuracy it is possible to indicate length on one of the component symbols to convey its relative perceptual weight.

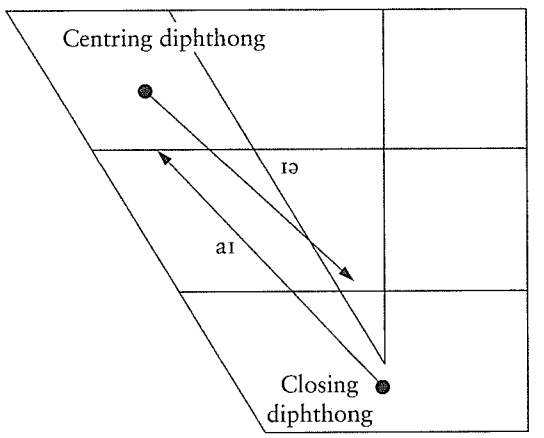

Figure 2.8.2 Closing and centring diphthongs

The durational structure of a diphthong also contributes to the distinction between a diphthong and a sequence of two vowels. If the glide component is quite short relative to targets of appreciable duration, the targets may be heard as two successive vowels. Readers may like to imagine a word *pawy* constructed from *paw* on the analogy of *handy* or *toey*. (Some speakers of English will be tempted to insert an *r* and to pronounce the word as *pory* but, for the sake of this illustration, they should resist the temptation.) Now by pronouncing *pawy* slower or faster and by adjusting the transition from one vowel to the next, it should be possible to vary the word from a distinctly bisyllabic word (*paw-ee*) to what sounds like a monosyllabic word containing a diphthong (*poy*). A simple exercise of this kind is useful in showing how we respond to variables in the flow of speech, but also points to the wider context of our judgements about syllables and other structural aspects of the organization of language which we shall treat further in the following chapter.

In summary, it is possible to produce an extraordinary range of different vowel sounds, although the repertoire of any one language will be confined within a system of relevant distinctions. Vowels can be described in terms of height and fronting (with suitable caution about the articulatory significance of these terms) and lip posture (rounded or unrounded); vowels are normally voiced but may in some languages have contrastive breathy voice or creak; vowels may also be nasalized (in opposition to normal oral vowels), tense (with advanced tongue root or widened pharynx) or long; and the presence of an onglide or offglide or the diphthongal combination of more than one vowel target adds a substantial range of auditorily distinct possibilities.

2.9 Consonantal sounds

In general, consonantal sounds show greater constriction of the vocal tract than vocalic sounds and have less prominence. Note that in English, as in many languages, a vowel can serve as an entire syllable (or word) as in *a*, *awe* or *I*, whereas consonants cannot. (It is certainly possible to produce some consonants without an accompanying vowel, as we do when we say *mm* or *sh*, but the structural organization of speech in most languages is such that vowels are normally central or nuclear in syllables, and consonants marginal or peripheral; see sections 3.1 and 3.11 below.)

There is a long tradition of drawing up inventories of symbols to represent all the various consonants of the world's languages, although the diversity of consonant articulation offers less scope for constructing a framework in the manner of the cardinal vowels (section 2.7 above). An early ambition of the International Phonetic Association (founded in 1886 and commonly referred to as the IPA) was to devise a universal phonetic alphabet, and the latest version of the Handbook of the International Phonetic Association (1999) lists and illustrates a set of symbols (including a version of the cardinal vowel system) that are widely known and used. We prefer here to put the emphasis on the ways in which consonants are articulated rather than on an inventory of symbols; but we still need to use symbols and will follow many of the IPA conventions, without overlooking other systems of representation that supplement or challenge the IPA scheme.

So far as their articulation is concerned, consonants can be described in terms of where the constriction is made, how it is made, and what kind of phonation supports it. Traditionally, especially in the IPA scheme, this is taken to mean that consonants can be displayed in a chart in which PLACES OF ARTICULATION are listed from left to right (from the front of the vocal tract to the back), and MANNERS OF ARTICULATION from top to bottom (from stops, with maximal constriction, through fricatives to various consonants produced with less constriction); in addition some consonants need to be specified as voiced or voiceless. Thus a typical chart of this kind has in the top row voiceless stops, beginning with bilabial [p] on the left and moving through various stops made in the oral cavity towards glottal stop at the extreme right. Below them are voiced stops likewise beginning with bilabial [b] on the left, and below them voiceless fricatives, and so on. The scheme reflects its European origins, as it omits, for example, ejectives and implosives (section 2.5 above). Moreover, while it is diagrammatically convenient to treat place and manner as single dimensions, this sometimes means that some features of articulation (such as the posture of the tongue) have to be ignored or dealt with outside the main chart or compressed into one of the two dimensions.

An early and influential critique of the IPA's style of phonetic description was Pike's (1943), which has significantly influenced later approaches to phonetic description, including our own. Pike was conscious of the need to broaden the range of languages on which phonetic generalities were being based, and he wanted to account fully and consistently for all the articulatory mechanisms that were available to humans. His survey is impressively thorough, and is often pursued in a spirit of exploring the limits of human noise-making rather than describing sounds known to occur in languages. Thus in addition to a comprehensive range of articulatory mechanisms, he mentions such exotic possibilities as producing an 'ingressive stop' by sucking the tongue tip from the bottom lip (1943, p. 101) and twisting the tongue lengthwise so that the tip is upside down against the teeth (1943, p. 122). Pike's system consequently allows for detail that cannot be readily justified in phonetic description, and its value lies in its challenge to traditional approaches and its influence on later work.

Some later descriptive frameworks have incorporated the results of modern instrumental research. The earliest and most comprehensive of these is Peterson and Shoup (1966a and b). This uses a primary articulatory description in terms of place and manner, but adds a series of secondary parameters to provide the necessary descriptive detail about airstream, airflow path and phonation mode. The authors also specify acoustic and physiological correlates for many of the dimensions of their system. Their place and manner specifications differ from traditional IPA practice in that they rank manner of articulation according to degree of stricture (from greatest to least) and specify both horizontal (lip to glottis) and vertical (tongue height) places of articulation. The system not only divides these dimensions more finely than is usual, but also places both

consonantal and vocalic sounds on a continuum using the one set of dimensions. Nevertheless, this treatment of vowels does not remove the difficulties of making accurate statements about tongue position in vowel sounds (section 2.7 above). In any event, despite Peterson and Shoup's logical and comprehensive approach, their system has not been widely used.

Among other contributions to general phonetic description, Catford (1977) is noteworthy: his objective is to account for all human articulatory possibilities – or 'anthropophonics' as he calls it, reviving a term used by Baudouin de Courtenay in the nineteenth century. Catford emphasizes the description of aerodynamic activity in articulatory processes, and offers more detailed categories for specifying articulatory locations (particularly on the tongue and lips) than are traditionally used. Catford also notes the inconsistency of using different descriptive systems for vowels and consonants, but concludes that the traditional method based on cardinal vowels remains the most practical. Other descriptive systems such as Jakobson et al. (1952), Jakobson and Halle (1956), Chomsky and Halle (1968) and Ladefoged (1971; 2006, pp. 268–76), which explicitly address the question of 'features' as the ultimate components of speech, will be discussed in chapter 10.

The outline which follows is based on the traditional dimensions of manner and place of articulation. To refine these rather constraining dimensions, individual articulatory processes which implement the dimensions are defined in some detail. The level of detail is obviously controversial – the framework presented here goes beyond the IPA scheme but stops short of Pike's attempt to capture everything that is physically possible. Ultimately a framework must be realistic, in the sense that it is adequate to account for the diversity of sounds actually encountered in languages without encompassing mere possibilities that are linguistically irrelevant. The symbols used are basically those of the IPA, with some extensions and minor changes. A summary chart of the symbols can be found in appendix 1.

2.10 Vocal tract place

The constriction that produces a particular consonantal sound is located at some point in the vocal tract. In traditional usage, a single value along the 'place of articulation' dimension defines *both* the area of the oral-pharyngeal vocal tract where the constriction is made *and* the part of the tongue used to form the constriction (if the tongue is the active articulator). In our scheme, VOCAL TRACT PLACE will refer only to location along the vocal tract. The posture of the tongue will be treated as a separate articulatory dimension, so that different tongue positions or gestures can be combined with different places of articulation. This approach is consistent both with Pike's method of description (1943) and with the spirit of much of the most recent work in phonetics.

The wall of the vocal tract, extending from the lips to the glottis, is a virtual continuum. There are some anatomical features – such as the teeth – which

constitute boundaries or areas, but much of the tract, and especially the roof of the mouth, does not divide naturally and obviously into regions. Thus the phonetic conventions governing the definition and labelling of articulatory areas owe rather more to observation of where sounds tend to be made than to anatomy. Points or places of articulation should therefore be understood as approximately demarcated regions rather than as specific points in the vocal tract. Partly for this reason, the labels of places of articulation are not entirely standardized, and we shall draw attention to some ambiguities. Figure 2.10.1 shows a mid-sagittal section of the supraglottal vocal tract indicating the articulatory locations described below. Most of the locations can be identified by looking in a mirror or by feeling inside the mouth with fingers or tongue.

LABIAL refers to the upper and lower lips. For description of languages we need to distinguish between BILABIAL articulation (both lips involved) and LABIO-DENTAL (lower lip and upper teeth). In English, [p] as in *pea*, [b] as in *bee* and [m] as in *me* are all bilabial; [f] as in *feel* and [v] as in *veal* are labio-dental.

DENTAL refers to the upper teeth. Apart from the involvement of the teeth in labio-dental articulation (above), various sounds can be made by the tongue against the teeth. Examples are English [θ] as in *thin* and [ð] as in *this*; [n] may also be dental when it precedes one of these, as in *month* or *ninth*. Some phoneticians distinguish between the edges of the upper teeth and the posterior faces of the upper teeth. It is true that the tongue may make contact with the teeth in different ways, but this can be largely explained by the way in which the tongue is used.

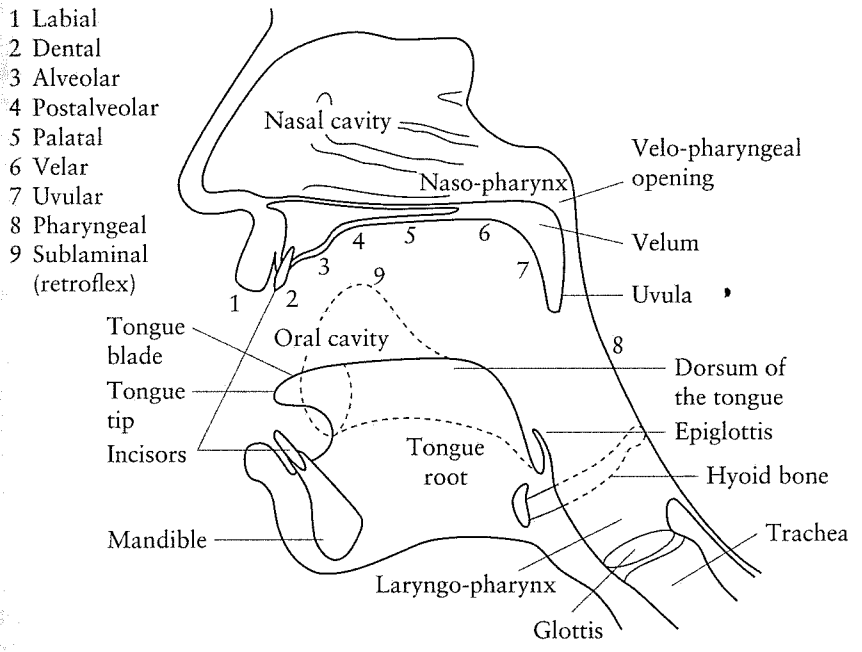

Figure 2.10.1 Mid-sagittal oral vocal tract showing major areas of articulation
Adapted from: Minifie et al. 1973, p. 173.

ALVEOLAR refers to the gum ridge or alveolum, the thick membranous covering on the bone structure which joins the tooth-bearing bone of the upper jaw and the vaulted or arched bone structure of the hard palate. The alveolum begins immediately behind the upper teeth and extends to the corrugations on the membranous covering of the posterior part of the tooth ridge structure. Some writers, such as Heffner (1964), have suggested that the anterior part of this alveolar ridge be considered separately as the 'gingival' (gum) region, but there is no linguistic justification for a distinction of this kind. Alveolar sounds in English include [t] as in *toe*, [d] as in *doe* and [n] as in *no*.

POSTALVEOLAR refers to the region from the corrugations on the tooth ridge where the roof of the mouth has a convex contour, to the start of the smooth surface of the hard palate where the roof of the mouth begins to become concave. For many (but not all) English speakers [ɹ] as in *ray* and [ʃ] as in *shy* are postalveolar consonants.

PALATAL refers to the region from the postalveolar area on the smooth surface of the hard palate to the start of the soft palate or velum. This is a larger articulatory area than those forward of it in the vocal tract; some phoneticians subdivide it into pre-palatal and palatal areas, but there do not appear to be any sounds in language that depend on such a distinction. The approximant consonant [j] as in *you* is palatal; most English speakers also advance [k] before a front vowel (as in *keep* or *king*) to such an extent that it is palatal.

There is some uncertainty among phoneticians about defining the alveolar, postalveolar and palatal regions. One problem here is that individuals differ somewhat in the anatomy of this region, another that significant differences in sound can be achieved by variation in tongue posture at the same (or nearly the same) point of articulation. The older terms 'palato-alveolar' and 'alveolo-palatal' illustrate the difficulties, for they purport to specify places of articulation, but imply particular tongue configurations as well. We allow for separate description of tongue configuration (section 2.11 below) and prefer to limit articulatory place to three areas (alveolar, postalveolar, palatal) which appear to be descriptively adequate when taken in conjunction with variation in tongue posture.

VELAR refers to the region extending from the start of the soft palate, or velum, back as far as the uvula. In English, [k] as in *core* and [g] as in *gore* are velar stops (but note that before front vowels, as in *keep* and *geese*, [k] and [g] are usually articulated much further forward, in the palatal region).

UVULAR refers to the short projection of soft tissue and muscle at the midline of the posterior termination of the velum. It is possible to close the back of the tongue against the uvula, as in the voiceless uvular stop [q] found in Arabic; uvular fricatives and uvular trills are also possible, and the *r*-sound of French and German is often articulated in this way.

The boundary between the hard palate and the velum is reasonably clear, since it lies at the posterior end of the bony extension of the upper jaw. There is no such boundary between the velum and the uvula. Many phoneticians avoid any specific definition of this division, or simply imply that velar articulations may occur on all parts of the soft palate except the uvula. Catford (1977), without supporting evidence, suggests that velar articulations occur only

in the anterior half of the region between the palatal–velar boundary and the uvula itself. The difficulty appears to be one of establishing the anterior limits of uvular articulation, which is affected by the nature of the articulatory activity concerned. Thus a uvular trill, which involves the uvular projection itself, occurs in the region of the posterior termination of the soft palate structure. A uvular stop, on the other hand, must be sufficiently far forward on the soft palate to ensure complete oral closure, and it does not require the uvular projection itself to serve as a dynamic articulator. The velar–uvular boundary is therefore not a sharp line, but rather an area slightly forward of the posterior termination of the velum.

PHARYNGEAL refers to the walls of the pharynx, including the root of the tongue. Pharyngeal consonants are not common but voiceless and voiced pharyngeal fricatives [ħ] and [ʕ] are found in several languages, of which the best known is Arabic.

GLOTTAL refers to the glottis, which plays a central role in phonation (section 2.6 above) but can also function as an articulator. Closure and release of the vocal folds, for example, can constitute a stop analogous to a bilabial or velar stop. This glottal stop is a consonant in languages as diverse as Arabic, Vietnamese and Hawaiian. English speakers tend not to hear the sound as a stop – or to count it as 'a catch in the throat' – and its occurrence in English as a nonstandard substitute for other stops is usually reckoned as omission of the correct stop (as in the kind of London pronunciation popularly represented as *pu*' for *put* or *ma'er* for *matter*).

2.11 Tongue position

Different parts of the tongue may be used in combination with the above places of articulation. The combinations are constrained in obvious ways: it is impossible, for example, to bring the back of the tongue into contact with the anterior regions of the mouth, at least not in such a way that one can usefully control an articulatory process. As there are no real landmarks on the tongue, the naming of points or areas on the tongue is in any case a matter of convention, influenced by those combinations of tongue position and place of articulation which prove to be functional in actual languages. (Figure 2.10.1 above includes a sagittal section of the tongue indicating such functional locations.)

APICAL refers to the tip or front edge of the tongue; LAMINAL to the anterior part of the upper surface of the tongue, otherwise known as the blade; and DORSAL to the region from the blade of the tongue to the root. The boundary between the laminal and dorsal areas is a matter of convention, but is generally defined as the region lying below the tooth ridge when the tongue is at rest. SUBLAMINAL is a term suggested by Catford (1977) to identify the anterior part of the undersurface of the tongue, corresponding to the blade.

English alveolar sounds such as [t], [d] and [n] are normally apical; dental versions of these sounds (as the dental [n̪] before [θ] in *month*) are also apical.

On the other hand many Australian Aboriginal languages (such as Aranda from Central Australia) have a dental stop which is laminal: the tongue is pushed forward so that the tip is down and the blade bunched against the back of the upper teeth. Use of the dorsal area of the tongue is often predictable from place of articulation – a velar or uvular constriction will inevitably involve the dorsal area – but it is also possible to bring the tongue forward in such a way that the dorsal area of the tongue approaches the palatal area of the roof of the mouth. The fronted English [k] in *keep* or *keen* is of this nature. Australian Aboriginal languages again provide a contrast, since they employ a palatal stop which is laminal rather than dorsal. The auditory difference is noteworthy: English speakers tend to hear the lamino-palatal stop as something like [t] immediately followed by [j] (thus combining the stoppage of a [t] with the lamino-palatal articulation of [j]) rather than as a fronted [k]. The term 'sub-laminal' is not widely employed but is useful in specifying tongue behaviour in so-called 'retroflex' consonants (found for example in some Australian Aboriginal languages as well as in many languages of India). In these consonants, the tongue may be curled up and back so that the undersurface of the front of the tongue makes contact with the roof of the mouth in the alveolar or postalveolar region.

Table 2.11.1 lists places of articulation in which various tongue positions are combined with various locations.

Table 2.11.1 Places of articulation for consonants

Name of place	*Articulators used*
Bilabial	Upper and lower lips (English *p*, *b*, *m*)
Labio-dental	Lower lip and edges of upper incisors (English *f*, *v*)
Apico-dental	Tongue tip and edges or backs of upper incisors (Spanish *t*, *d*, English *th* in *thin*)
Lamino-dental	Tongue blade and edges or backs of upper incisors (*th* in Australian Aboriginal languages)
Apico-alveolar	Tongue tip and alveolar region (English *t*, *d*)
Lamino-alveolar	Tongue blade and alveolar region
Apico-postalveolar	Tongue tip and postalveolar region (southern British English *r* in *trip*, *drip*)
Lamino-postalveolar	Tongue blade and postalveolar region (English *sh* as in *ship* may be apico-postalveolar or lamino-postalveolar depending on the speaker)
Sublamino-postalveolar	Tongue undersurface and postalveolar region (as in 'retroflex' sounds of Hindi or Urdu)
Apico-palatal	Tongue tip and palatal region
Lamino-palatal	Tongue blade and palatal region (English *y*)
Velar	Tongue body and soft palate (English *k*)
Uvular	Tongue body and uvula/soft palate (*r* in some varieties of French and German)
Pharyngeal	Pharynx walls
Glottal	Glottis (vocal folds)

2.12 Manner of articulation

Manner of articulation covers both the degree or extent of a constriction and the way in which the constriction is formed in the vocal tract. Thus a category such as 'stop' implies both blockage of the airstream (total constriction) and a movement to create and then release the blockage (dynamic articulation). On the other hand, 'fricative' implies a lesser constriction and a kind of articulation which could, in principle, be prolonged as a steady state (stable articulation). In traditional descriptions (such as those following the IPA conventions), manner of articulation can sometimes also include a specification of constriction shape, for example in descriptions such as 'lateral fricative', where 'lateral' refers to tongue configuration against the roof of the mouth. But since it is possible to vary the shape of a constriction independently of the other aspects of manner of articulation, we deal with shape separately as STRICTURE (section 2.13 below).

Like most other articulatory variables, consonantal constriction is a continuum. It ranges from total closure of the vocal tract to fully open, vowel-like articulation. For linguistic description, a three-way distinction of stoppage, fricative articulation and a more open vowel-like articulation appears adequate. In English, [b] requires stoppage, [v] a fricative constriction, and [w] a still wider constriction. The distinction rests primarily on the effect of each degree of constriction on the airflow, and secondarily on the kind of articulatory manoeuvre that produces the constriction.

This relatively simple classification is complicated by the further distinction between dynamic and stable articulations. A stop is necessarily dynamic: it is characterized by the actions of forming and releasing the stoppage. Note that one cannot greatly prolong a stop such as [b] or [p] other than by maintaining the closure (in which case no sound is heard during the closure) or repeating the actions of closing and releasing the articulators. Readers may like to verify this by experimenting with a word such as *happen*. It is possible to hold the [p] closure for some time, but no sound is then heard; and it is possible to repeat the [p] a number of times as if stammering over the consonant; but there is no way to prolong the sound of a single [p]. On the other hand, the more open constrictions of sounds such as [v] and [w] are stable in the sense that they can be prolonged in a more or less steady state. It is possible, for instance, to hold the [v] in *ever* as long as one's breath lasts. The [w] in, say, *owing* can be similarly prolonged – and in keeping with its vowel-like character will sound much like a lengthened [u] vowel. (In normal running speech, of course, stable articulations are very brief and may not necessarily be perceived or identified as 'steady states'.)

Other manners of articulation extend this repertoire. First, nasal consonants are in one sense stops, for the airflow is blocked at some point in the oral cavity; but since the velum is lowered to allow airflow through the nasal cavity, nasal consonants can be prolonged (and commonly are in what we call 'humming a tune'). Nasals are therefore classified not as stops but as a separate

manner of articulation. Secondly, there are other kinds of dynamic articulation besides stops. Accordingly, we recognize seven manners of articulation: STOP, FRICATIVE and the vowel-like APPROXIMANT; NASAL; and three additional dynamic manners, FLAP, TAP and TRILL. These terms are standard, except that approximant consonants have been variously defined and labelled: terms such as GLIDE, FRICTIONLESS CONTINUANT, ORAL RESONANT and SEMIVOWEL are sometimes used for one or more kinds of approximant. And the term OBSTRUENT is commonly used to include both stops and fricatives.

A STOP is produced by the formation and rapid release of a complete closure at any point in the vocal tract from the glottis to the lips. The velum is raised to prevent airflow through the nasal cavity, and the oral airflow is thus interrupted. The durations of the phases of a stop are partly conditioned by phonetic context and therefore variable: the stoppage itself may last from 40 to 150 ms, and the closure and release phases may each last between 20 and 80 ms. The release of a stop is particularly complex, as several factors are relevant. First, the nature of the airflow during release is largely dependent on the nature of the glottal airflow (defined by phonation, section 2.6 above). Secondly, timing is also significant, as the moment of release need not coincide exactly with other articulatory gestures (such as the start or finish of voicing). Thirdly, the stoppage itself creates a change in pressure. If the airstream is egressive (whether pulmonic or glottalic), air pressure will build up in the oral cavity behind the occlusion; if the airstream is ingressive (whether glottalic or velaric), intra-oral air pressure is likely to be reduced during the occlusion. Egressive pulmonic stops are by far the most common type of stop and are sometimes identified by the label PLOSIVE.

In a typical voiced plosive, there must be airflow through the glottis to generate the voiced phonation. But the very nature of a stop is that airflow is blocked somewhere in the vocal tract. As pressure builds up behind this blockage, it will approach the level of subglottal pressure generating airflow through the glottis, eventually to the point where phonation cannot be sustained. At this point, of course, the stop is no longer a voiced stop, but voiceless. There are various linguistic responses to this aerodynamic problem. One is that voiced stops in many languages often are partially devoiced. In English, for instance, it is common for voicing to tail off in a word-final voiced stop (as in *rib* or *rid* or *rig*). It is also noteworthy that the occlusion phase is often shorter in voiced stops than in voiceless, so that the vocal apparatus is, so to speak, not put to the test of maintaining voicing for any length of time. And it is also possible to enlarge the space between the glottis and the point of stoppage during the occlusion, by lowering the larynx and distending the pharyngeal walls to increase cavity volume. In this way, intra-oral air pressure is not allowed to build up as quickly, and a pressure drop across the glottis can be maintained (Ohala 1978).

At the release of a stop, there is a very short sharp pulse of turbulent airflow through the (momentarily) narrow aperture of the parting articulators. During this pulse – known as the 'release burst' – the peak airflow rate can exceed 1.5 litres per second. After the release, the articulators move rapidly to the next required position. The actual phonetic quality of the release burst and

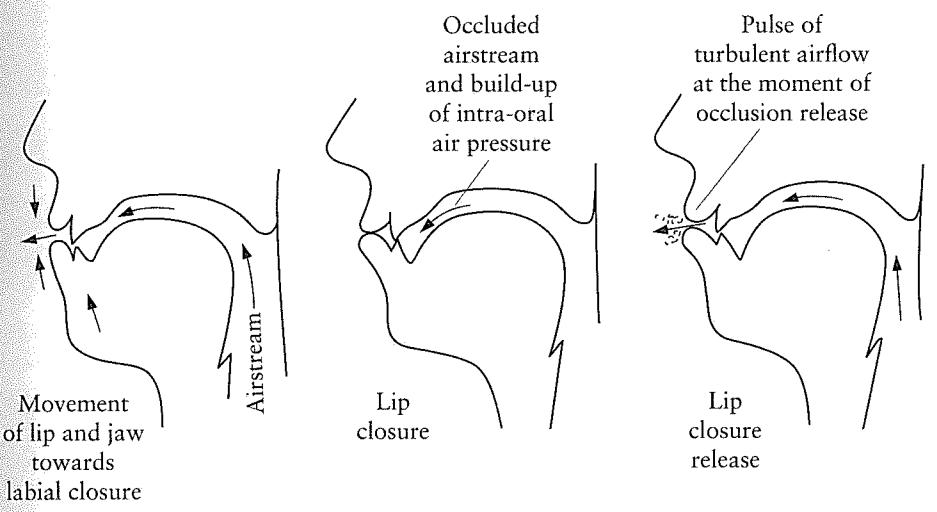

Figure 2.12.1 Phases of a bilabial plosive

what follows it is dependent first on the place of articulation, and secondly on the phonation mode at the time. Figure 2.12.1 shows the formation, occlusion and release of a bilabial stop with a normal pulmonic egressive airstream.

Readers can check some of these observations by producing an [a:] sound (*aah*) and then closing and releasing the lips at intervals to produce [a:ba:ba . . .]. Prolonging a closure will illustrate the devoicing which occurs as transglottal pressure falls; and a deliberate change of phonation mode from voiced [b] to voiceless [p] during the occlusion phase should enable the reader to sense the different demands of voiced and voiceless stops. Many languages exploit the distinction of voiced and voiceless stops, but with considerable variation in the way the distinction is realized or implemented (especially in relation to the timing of voicing, section 2.16 below). Only the glottal stop cannot be voiced, as the glottal occlusion obviously rules out any possibility of voiced phonation.

FRICATIVE is a potentially stable articulation produced by a constriction in the vocal tract that is narrow enough to create turbulent airflow. The noise of this turbulence (modified by the effects of the vocal tract shape) gives many fricative sounds a characteristic hissing or sibilant quality. Figure 2.12.2 illustrates the airflow pattern of the fricative [s].

The factors that make one fricative sound different from another are place of articulation, the shape of the constriction, and the aerodynamic forces of the airstream. Additionally, in the case of dental, alveolar and postalveolar fricatives, the front (incisor) teeth contribute to phonetic quality, since they deflect the airflow coming from the constriction, producing some additional turbulence.

There is a balance between the cross-sectional area of a fricative constriction and the rate of airflow through the constriction. The constriction must be relatively small to generate turbulence, but the air must also flow rapidly enough to exceed the threshold at which smooth or laminar airflow becomes turbulent.

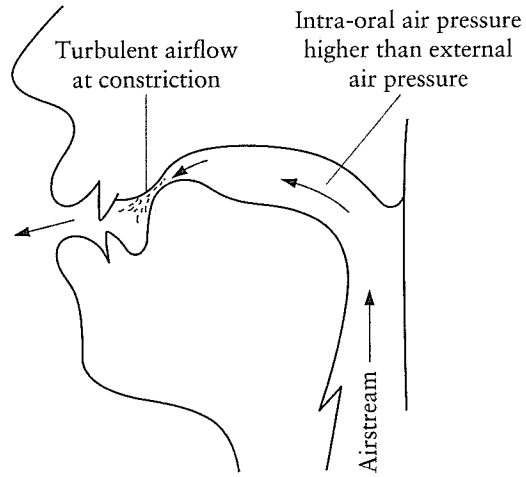

Figure 2.12.2 Articulation of a voiceless alveolar fricative [s]

If the constriction area is enlarged, then the flow rate must be higher to achieve turbulence. Studies of fricative aerodynamics by Hixon (1966), Stevens (1972b), Warren (1976), Catford (1977) and Shadle (1991) indicate that fricative constrictions of up to around 30 mm^2 can produce turbulence, provided airflow is high enough. Flow rates vary between about 30 and 300 cm^3 per second. Models of fricative articulation as well as empirical data suggest that there is considerable room for variation and that individual speakers may have different articulatory habits. Air pressure in the oral cavity behind the fricative constriction seems to depend on the rate of airflow and size of constriction, and is thus not a primary factor in determining the nature of a fricative. Reasonably typical pressure and flow patterns of a voiceless alveolar fricative are shown in figure 2.12.3 (Clark et al. 1982).

The peaks of airflow at the start and end of the fricative in figure 2.12.3 reflect the relatively large area of constriction during the formation and release of the fricative. Between the peaks, airflow falls to a minimum, corresponding more or less to the period of maximum constriction. The peak of intra-oral air pressure, as might be expected, also corresponds to the period of maximum constriction and maximum airflow resistance. Airflow rates are of course also affected by phonation. In a voiceless fricative, there is negligible resistance to airflow at the glottis, and airflow will be higher than in a voiced fricative. In voiced fricatives, not only is airflow resistance higher at the glottis, the flow is also interrupted at the rate of vocal fold vibration. This intermittent effect on the turbulence at the fricative constriction is largely responsible for the voiced quality of the sound (Klatt et al. 1968, Scully 1979).

To check the effects of fricative constriction on airflow, readers may like to produce a continuous [s] sound and then pull the tongue down from the alveolar ridge. Airflow will increase rapidly and fricative noise will suddenly cease as airflow through the constriction switches from turbulent to smooth or laminar flow, as shown in figure 2.12.3.

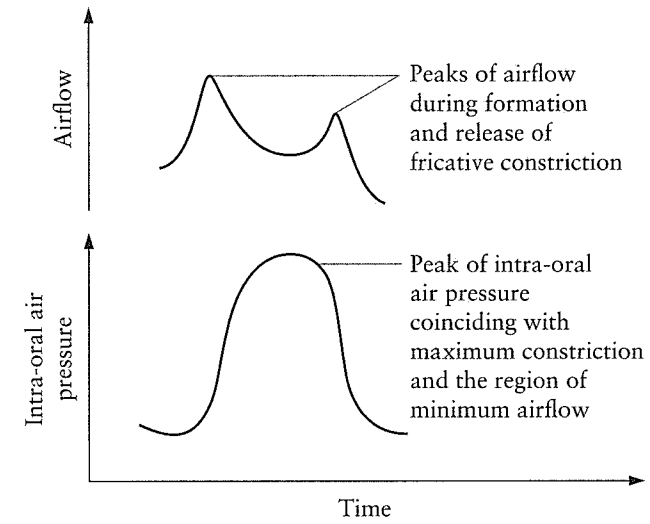

Figure 2.12.3 Pressure and airflow patterns for a voiceless alveolar fricative [s]

An APPROXIMANT is a potentially stable articulation in which the constriction is normally greater than in a vowel, but not great enough to produce turbulence at the point of constriction. Following Ladefoged, the term covers the traditional categories of 'frictionless continuant', 'semivowel' and 'oral resonant'. For Abercrombie (1967) and Catford (1977), approximant is a narrower class of sounds, excluding laterals. In English, the consonants [l], [w] and [j] as heard at the beginning of *law*, *war* and *your* are all approximants in the wider sense. The initial consonant in *raw* is also an approximant, at least for those speakers of English who do not trill or flap the *r*.

The maximum degree of constriction in an approximant is defined by the onset of turbulent airflow: if constriction is great enough to create turbulent airflow, the sound is a fricative, not an approximant. The minimum degree of constriction is less clear-cut. Even if we take it that most approximants have greater constriction than most vowels, we may find little or no articulatory difference between high (close) vowels such as [i] and [u] and their 'semivowel' counterparts [j] and [w]. While an articulatory distinction can be made, the difference often has more to do with syllabic organization than with articulation of the sounds themselves (see chapter 3 below, especially sections 3.11 and 3.13).

Approximants are normally voiced, and by definition cannot have turbulent excitation at the point of constriction. In theory, it is possible to produce voiceless approximants, but they require a noise source, such as turbulence at the glottis created by whisper phonation or some more generally distributed turbulence in the vocal tract created by high volume airflow. In practice, it is difficult to distinguish between a voiceless approximant and a voiceless fricative at the same place of articulation. Thus in English, approximants may be devoiced following voiceless consonants, for instance the [w] in *twin* or *twelve*. This voiceless approximant is in effect a voiceless bilabial fricative with lip rounding –

or, more pertinently, it makes no difference to the English sound system whether the sound is regarded as an approximant or a fricative. And there is no evidence that any language in the world makes such a distinction crucial.

NASAL consonants have a stoppage at some point in the oral cavity; at the same time, the velum is lowered to allow airflow through the nasal cavity. The sounds are therefore perceived as potentially stable and continuous rather than as stops in the true sense. Common nasal consonants are [m] and [n] (as in English *more* and *nor*). English speakers should have no difficulty in verifying that a nasal consonant can be prolonged, as in a thoughtful *mmm*.

FLAP and TAP are dynamic articulations in which there is a very brief occlusion in the vocal tract. The terms are sometimes used synonymously, but it is possible to distinguish two kinds of action: in a flap, one articulator strikes another in passing, not so much to create a brief closure but more as the incidental effect of the articulatory gesture; in a tap, there is a single deliberate movement to create a closure, tantamount to a very short stop.

The most common flaps are ones in which the tongue strikes the alveolar ridge in passing. Many speakers of English use a flapped *r* in words such as *three* and *throw*, where the tip of the tongue strikes the alveolar ridge on its way from the dental position to a more retracted position for the following vowel. Some languages, including Hindi and the Central Australian language Warlpiri, have a flapped *r* articulated somewhat differently: the tongue tip may be curled back towards the palate and may then strike the posterior part of the alveolar ridge as it moves down towards its neutral or rest position. Ladefoged (2006, p. 172) also reports a labio-dental flap (from Margi, a language of Nigeria) in which the lower lip is drawn in and then allowed to flap against the upper teeth as it returns forward. The most commonly cited instance of a tap is from some varieties of English: some speakers, especially Americans but also younger Australians, pronounce the medial [t] in words such as *better* and *matter* as a tap. The pronunciation often strikes other speakers as converting a [t] into a [d] – and indeed a tap against the alveolar ridge is in one sense simply a very short [d]. But those who use a flapped *r* in English, such as Scottish speakers, may hear the sound as closer to the flap than to a stop. Moreover, it should be noted that speakers who use the tap do not normally confuse the tap with [d], and *matter* can still be distinguished from *madder* (although the length of the preceding vowel rather than the nature of the tap or stop may play the crucial role here).

TRILL is a dynamic articulation produced by vibration of an articulator. The articulatory setting is such that the articulator is not deliberately moved but vibrates as a consequence of the egressive airstream passing by it. The airstream is repeatedly interrupted at a rapid rate, rather in the way that voiced phonation is produced by vibration of the vocal folds. The most common trills use the tongue tip (held close to the alveolar ridge) or the uvula (by bringing the dorsum of the tongue into light contact with it). A trill is a series of vibrations and is described as a dynamic articulation because no single vibration can be lengthened significantly. But the trill itself can of course be lengthened by repeating the vibrations, in principle indefinitely, as long as the airflow lasts. But in normal speech, trills tend to be short and to use rather few vibrations.

Most readers will be familiar with the alveolar trill as a 'trilled r', even if they do not normally use it in their own speech. In fact, outside English many people will consider an alveolar trill to be the common or normal way of articulating an [r]. Speakers of Italian, Spanish and Indonesian, for example, readily trill the [r], particularly if speaking emphatically or clearly. (Other articulations, including tap or flap mechanisms, may be used in these languages; see Lindau 1985 for general discussion of 'r-sounds'.) The alveolar trill is not common in English, except in Scottish English, where it may be the normal articulation. Some speakers of French and German use a uvular trill as their 'r-sound' – but again other articulations may also be used, including a uvular fricative or approximant and, especially in stage pronunciation or in some rural dialects, an alveolar trill. Thus in many languages – English, French, German, Italian and Indonesian among others – there is only one 'r-sound' and variations in the pronunciation of it are associated with regional, social or stylistic differentiation. On the other hand, Spanish distinguishes between a flap (as in *pero*, 'but') and a trill (as in *perro* 'dog') while Warlpiri, from Central Australia, has three kinds of *r* – an alveolar flap or trill, a retroflex flap, and an approximant (similar to the *r* used by most English speakers).

2.13 Stricture

Stricture refers to the shape of a constriction. For many sounds, stricture is either irrelevant or determined by other aspects of the articulatory process. In a nasal consonant such as [n], for instance, the tongue makes a closure against the alveolar ridge while air flows through the nasal cavity, which offers no option of varying constriction shape; or in a flap or trill, the stricture will simply be a consequence of the vibratory movement of one articulator against another. In some articulations, however, the posture of the tongue can make appreciable differences in the shape of the constriction and the resulting quality of sound. In the fricative [s], as in *saw*, the tongue is grooved along its length in a way that contrasts with the flatter tongue shape of the fricative [θ] in *thaw* or the approximant [ɹ] in *raw*; while in a lateral sound such as the [l] in *law*, the tongue makes contact against the alveolar ridge but is lowered at one or both sides so that air flows through relatively freely. Figure 2.13.1 shows schematic sections for the three stricture types CENTRAL, GROOVED and LATERAL.

CENTRAL can be taken in a general sense to be the neutral value of stricture, applying to any constriction in which the tongue does not adopt a distinctively grooved or lateral posture. The term is more narrowly justified in instances where airflow along the centre of the vocal tract is in direct contrast with grooved or lateral stricture. Compare the initial consonants of English *trip* and *chip*. In many varieties of English the two words are auditorily and articulatorily quite similar. In the first word, the initial [t] is followed by a voiceless fricative (a devoiced and fricative counterpart of the common approximant

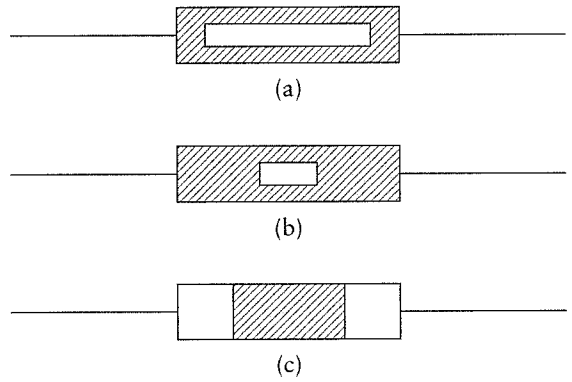

Figure 2.13.1 Stricture types: (a) central; (b) grooved; (c) lateral

value of English *r*); in the second word, [t] is followed by a grooved fricative, more or less identical with the one written as *sh* in *ship*. (Readers may like to pause over these examples. Comparison of *ship* and *chip* and an attempt at pronouncing the beginning of *chip* very slowly should highlight the nature of the fricative component. It may then be possible to compare *trip* and *chip* to verify that the shape of the tongue is different in the two fricatives, although details of the articulation will be by no means identical for all speakers of English. It may also be of incidental interest that young children sometimes have difficulty in grasping the spelling conventions and may produce written efforts such as *chrip* for *trip*.)

GROOVED refers to the tongue grooving already mentioned above, which yields a constriction of very narrow cross-sectional area along the vocal tract. Grooving is in fact common in fricatives which use the tongue tip or blade as the constricting articulator in the alveolar or postalveolar region. The grooved fricative [s] is found in many languages (but not in all – most Australian Aboriginal languages do not have it) and is far more common than flatter or more central fricatives of the kind heard in *trip*. Other quite frequent grooved fricatives are the voiced counterpart of [s], namely [z], and the voiceless postalveolar [ʃ] (as in English *ship*). English also has a voiced partner for [ʃ], namely [ʒ], heard in the medial position of words such as *measure* and *fusion*.

LATERAL refers to constrictions in which the airstream is diverted from the centre of the oral tract and flows to one or both sides. An alveolar lateral approximant [l] (English *law*) is widespread, but other points of articulation are possible (for instance the lamino-palatal lateral approximant [ʎ], represented as *gl* in Italian and as *ll* in standard European Spanish). A voiceless version of [l] is also possible, although as with other approximants, voicelessness is likely to imply fricative articulation. Thus the Welsh sound represented as *ll* (as in *llyn* 'lake' or *llan* 'church') is sometimes referred to simply as 'voiceless *l*' but probably is more strictly a voiceless lateral fricative [ɬ], having turbulent airflow through the lateral stricture. If airflow is high enough, turbulence can be achieved in a voiced lateral, yielding a voiced lateral fricative [ɮ] (found in Zulu). Indeed, lateral stricture can combine with various other articulatory processes, even with a click mechanism (as in the sound known to most English speakers as the noise to urge a horse forward, section 2.5 above).

2.14 Force

A distinction can be made between FORTIS articulation, relatively strong or forceful overall articulation, and LENIS or weak articulation. Our use of the terms follows Pike, who says that fortis articulation 'entails strong, tense movements . . . relative to a norm assumed for all sounds' (1943, p. 128). Fortis articulation is probably mainly a matter of greater subglottal pressure (and the term 'heightened subglottal pressure' has been used in some descriptions) but higher airflow and stronger and more definite supraglottal articulatory gestures are likely to accompany an increase in subglottal pressure.

Fortis articulation is sometimes described as 'tense' articulation: given the divergent uses of this term, we restrict it to vocalic sounds (and even there leave a question mark against it). Fortis articulation is also sometimes used to account for what we call 'aspiration' of stops: we deal with this as a matter of timing (section 2.16 below). We retain the fortis/lenis distinction because it is evident that a variable of force can be exploited in the articulation of certain consonants, alongside other variables such as phonation and timing. In Dutch, for example, the voiceless [p] heard at the beginning of words such as *pen* 'pen' and *pan* 'pan' is not aspirated but is typically articulated quite forcefully. (A similar kind of articulation may be heard from some English speakers from northern England; more commonly, English *pen* and *pan* will have initial stops that are voiceless, fortis *and* aspirated, section 2.16 below.) But the exploitation of the processes and the precise way in which they are integrated for an overall distinctive effect varies from language to language. We urge caution in drawing conclusions about the universal nature of articulation from detailed phonetic investigation of sounds in a single language. (See also our comments on TENSE and LAX at the end of section 2.7 above, the description of the STOP manner of articulation in section 2.12 above, and discussion of VOICE ONSET in section 2.16 below.)

2.15 Length

Like vowels (sections 2.7 and 2.8), consonants may be SHORT or LONG. Virtually any consonant can be made relatively longer or shorter, although in some cases the longer and shorter versions may count as different manners of articulation. Thus, one view of a flap is that it is a minimal trill, while a trill can be regarded as a series of flaps. Similarly, a tap may be considered a very short stop. Even

voiced plosives, in which voicing will cease during prolonged closure, can be lengthened sufficiently to be noticeably different in duration from a shorter version.

Languages which distinguish long and short consonants of various kinds include Italian and Finnish. The spelling system of both languages uses double letters to represent the lengthened consonants, e.g. Italian *notte* 'night', *canne* 'canes' (versus *note* 'notes', *cane* 'dog'). Lengthened consonants are often treated as the uninterrupted succession of two identical short consonants (as implied by the Italian spellings), in which case they may be called GEMINATES.

2.16 Voice onset

This variable refers to the timing of the start of voiced phonation relative to the supraglottal activity. It is mainly relevant to stops, and we consider three simple possibilities. If voiced phonation begins at or before the formation of the occlusion, the stop will count as FULLY VOICED. If there is no voicing during the occlusion but voice onset is virtually simultaneous with the release of the occlusion, the stop can be regarded as VOICELESS. If voice onset is significantly delayed beyond the release of the occlusion, the release will be ASPIRATED because of the unobstructed flow of air through the abducted folds. The timing is obviously a continuum and can be varied between these three options, but the three are more than sufficient to account for what happens in many languages.

In many languages, the timing of voice onset is actually redundant with respect to phonation values. In English and German, for example, word-initial voiceless plosives [p] [t] [k] are generally aspirated, i.e. there is some delay in voice onset after the release. In these same two languages word-initial voiced plosives [b] [d] [g] may not be fully voiced, i.e. voice onset may be somewhat later than the start of occlusion. Under these circumstances, the voicing of voiced plosives may not be very prominent and the aspiration of voiceless plosives may be the major factor in identifying their voicelessness. Descriptions of the languages nevertheless refer to 'voiced' and 'voiceless' plosives on the assumption that voice onset is predictably delayed. On the other hand, there are languages in which voice onset is not significantly delayed, for example French and Dutch. In these languages, voiceless plosives are not normally aspirated, i.e. voice onset virtually coincides with the release of occlusion. As might be expected, the voiced plosives of these languages are fully voiced, i.e. voicing begins at or before the start of occlusion.

Languages are not restricted to a two-way distinction among stops. There are languages which distinguish three kinds of plosive, generally referred to as 'voiceless aspirated', 'voiceless unaspirated' and 'voiced'. These languages include a number of East Asian languages such as Thai and Burmese. Here voice onset cannot be regarded as a redundant or secondary factor, and the three options outlined above are relevant, namely: FULLY VOICED, where voice onset or voice onset time (VOT) is at or near the beginning of the occlusion phase of the stop (often also referred to as VOT lead or 'negative' VOT);

UNASPIRATED, where voice onset occurs as the occlusion is released (often referred to as coincident VOT); and ASPIRATED, where voice onset is appreciably later than the release of the occlusion (often referred to as VOT lag or 'positive' VOT).

Finer differentiation is certainly possible, by the use of values such as PARTIALLY VOICED (for a word-initial plosive in which voicing begins during occlusion or for a word-final plosive in which voicing tails off well before release) or WEAKLY ASPIRATED and STRONGLY ASPIRATED (to distinguish between plosives with moderate delay in voice onset and those with considerable delay – which seems to be necessary in Korean, according to Kim 1965).

It should also be noted that plosives in most languages behave differently in different environments. In English, for example, voiceless plosives can be said to be noticeably aspirated; but this is most strikingly true of plosives standing word-initial before a vowel (as in *pea*, *tea*, *key*), less evident where a plosive stands between vowels (as in *happy*, *natty*, *lackey*) and generally not true at all of plosives following *s* (as in *spare*, *stare*, *scare*). In short, what appears to be a consistent distinction may be quite variable, and the precise cues that differentiate, say, *tie* from *die* may not be at all the same as those that distinguish between *matter* and *madder* or *mat* and *mad*.

Indeed, since variation in voice onset co-occurs with selection of various phonation types, of fortis or lenis articulation, and so on, a rich diversity of plosive types can be found in the world's languages. The voiced aspirated plosives of many South Asian languages, such as Hindi and Gujarati, for example, exploit breathy voice combined with some delay in the onset of normal voicing. Rarely, voicing delay can be observed in sounds other than plosives: a few languages have voiceless aspirated fricatives such as the [s^h] of Burmese.

The classic study of voice onset is Lisker and Abramson (1964). Ladefoged (1971) provides a useful survey. He distinguishes five values of voice onset for stops and fricatives, as follows:

1 voicing throughout (voiced);
2 voicing in part;
3 voicing starts immediately after (voiceless unaspirated);
4 voicing starts shortly after (slightly aspirated);
5 voicing starts considerably later (aspirated).

He also notes (1971, p. 20) that the five are merely points on a continuum and that no language seems to contrast more than three points.

See Lindblom and Maddieson (1988) for a useful discussion of phonetic universals in consonant systems, and Ladefoged and Maddieson (1996) and Ladefoged (2001) for a more general survey of consonant and vowel sounds in the world's languages.

Exercises

1 Check that you understand the meaning of each of the following terms used in describing speech sounds.

a. alveolar
b. approximant
c. aspirated
d. fricative
e. labio-dental
f. laminal
g. lateral stricture
h. lenis
i. palatal
j. velar

2 What are airstream mechanisms?
3 Explain briefly how whisper and breathy voice differ from normal voicing and from each other.
4 Explain the origin and use of cardinal vowels.
5 What is a glottal stop? Give examples of it from as many languages as possible.
6 How is the timing of voice onset significant in some languages?
7 Describe how each of the following sounds is made.

a. [s]
b. [m]
c. [g]
d. trilled [r]
e. [u]

8 How many vowels and how many syllables are there in each of the following words?

coward, crowd, groan, grown, higher, hire, line, lion, hour, our

9 What is stricture in the description of consonants?

3 Units of Speech

Introduction

This chapter brings us to a consideration of speech sounds as units. The chapter begins (3.1) with a discussion of what actually constitutes a unit of spoken language. It then introduces the concept of complex articulations, articulations in which gestures or settings overlap or are combined to produce what appear to be unitary sounds (3.2).

Specific complex articulations are then described:

- nasalization (3.3)
- labialization (3.4)
- palatalization (3.5)
- velarization and pharyngealization (3.6)
- affrication (3.7)
- double articulation, combining two places of articulation (3.8)
- vowel retroflexion (3.9)
- diphthongization (3.10).

This survey of complex articulation raises several questions about the distinction between consonants and vowels and about the ways in which languages organize syllabic structure. These questions are explained and addressed under the following headings:

- syllabicity (3.11)
- segmentation and structure (3.12)
- diphthongs and related phenomena (3.13).

The chapter ends with an account of how linguists have conventionally 'interpreted' the flow of speech as a linear structure appropriate to the language being analysed (3.14).