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).

3.1 Identifying the units of speech

In describing language we need to refer to units of language. The fundamental reason for this is not just that it is traditional and convenient to refer to sounds and words and syllables and other such elements; it is that language itself depends on discrete and finite options. The point is a general one, not limited to phonology. Human beings can, for example, distinguish a vast number of colours, and frequently do make very fine distinctions when matching paints or dyes. Indeed, colour is a continuum (or in our usual experience of it, at least three intersecting continua, technically referred to as hue, saturation and brightness). These continua are infinitely divisible down to the limits of our perception, yet English represents only limited choices: a handful of colour terms, a few adjectives that can be combined with them (as in 'bright red', 'deep blue') and the general devices of English grammar that allow for relative judgements (as in 'this dress is a darker red than that one' or 'this is not very blue at all, more a grey'). Of course other languages have other words and other mechanisms, and they may divide the continua in quite different ways, but in principle they will all use limited options.

The point applies fully to speech and hearing. Pitch height is a continuum along which we can and do make fine distinctions (as in music, where we hear small variations as sharp or flat notes) but our linguistic systems exploit simple and relative contrasts, such as 'high' versus 'low', or 'high' versus 'mid' versus 'low' (Gussenhoven 2004, p. 27). Vowels likewise (section 2.7 above) fall within a space that can be very finely subdivided, yet most languages exploit rather few contrasts with often no more than half a dozen contrasting vowel qualities.

Thus units point to the systemic nature of language. Wherever a finite number of elements forms a set of contrasting options, we may speak of a system (or a subsystem within a larger system). But the term also points to structuring, or linear arrangement, for a unit characteristically belongs to a particular level of linguistic organization. On one phonological level we may speak of speech sounds as units, on another level we may recognize syllables as units, and on yet another, phonological words or phrases. On one level, the English word *conduct* is a sequence of seven sounds (which happen to be reasonably well indicated by the seven letters of the written form); on another level, it is a sequence of two syllables; on another, it is a single unit, a word.

These units are not ends in themselves but are justified by their descriptive validity. If we want to explain patterns of English stress, for example, we must recognize syllables as relevant units. Stress is a relative property and it must be defined over syllables – we hear a syllable as stressed because it stands out against something which counts as unstressed. In other words, a syllable is not stressed or unstressed in absolute terms, but is more stressed or less stressed than a neighbouring syllable or some other point of reference. Hence when we say, as part of the description of English, that English words such as *conduct*, *insert* and *produce* each have two patterns of stress (one signalling a verb, the other a noun) we need to refer not to properties of sounds but to relative values of syllables.

Moreover, the validity of units such as syllables can frequently be justified by appeal to the behaviour and judgements of native speakers of a language. English speakers, for example, can normally say how many syllables a word has, and make implicit use of this sense of syllabic structure when they construct verse or put words to a tune.

The units of a language are thus determined in relation to the system and structure of that particular language, although it is reasonable to assume that most languages will have comparable elements (such as sounds and syllables). But even if we can establish what the units of a language are, a further challenge arises when we try to determine the boundaries of those units – when we try to segment connected speech into a chain of sounds or syllables. In chapter 2 we assumed that discrete sound units could indeed be identified in any language (section 2.4), but we could not entirely avoid certain difficulties. It was apparent, for example, in the discussion of long vowels and diphthongs (section 2.8) that the distinction between a single segment and a succession of two segments was by no means straightforward.

Readers may have accepted fairly readily the assumption that speech consists of a sequence of individual units – perhaps too readily, and probably because of our alphabetic writing system, rather than because of any persuasion about the nature of spoken language. The custom of using sequences of discrete letters to write a language such as English does offer an immediate analogy for the segmentation of speech, and the conventional spelling of, say, *conduct* may make it seem obvious that the word consists of seven segmental sounds. But familiarity with alphabetic writing can be highly misleading, for the parallel between writing and speech is not exact. There are obvious examples in English spelling, where 'silent' letters (as in *psalm* or *knight*) and 'digraphs' (two letters for a single sound, e.g. *th*, *sh*) complicate the relationship between letters and sounds. Moreover, although written English is conventionally organized into various units (by spaces between words, punctuation marks between clauses or sentences, etc.), the units of written English do not necessarily match those of spoken English. There is, for instance, no regular indication of syllables in written English. Even when words are hyphenated at the ends of lines the hyphens often do not coincide with the boundaries of spoken syllables. Even more importantly, the organization of speech is *in principle* independent of orthography: languages without regular written form still have phonological system and structure, and persons who cannot read or write still use organized speech. It may well be true that knowledge of a writing system facilitates a certain analytical awareness of segments and structure, but the absence of written expression does not rob spoken language of its systematic nature (Halliday 1985b, esp. ch. 7).

The experience of reciting and listing the letters of an alphabet encourages the notion that segmental sounds have an independent existence, and literate speakers therefore tend to overlook the analytical process that is involved in segmentation and even to assume that individual letters prove the existence of individual sounds. But there are well-established nonalphabetic writing systems,

such as the Chinese character system and the Japanese Hiragana and Katakana syllabaries, in which symbols stand for entire syllables (or even words), and this is sufficient evidence that orthography is not bound to represent discrete sounds. Moreover, conventional spelling too easily persuades us to ignore the continuous nature of speech. Thus the natural transition between an [n] and a [z] in words such as *frenzy* and *bronze* could easily be taken as the consonant [d], but it is unlikely to be perceived as such by English speakers who are aware of the spelling; on the other hand, the [d] in, say, *friends* or *fronds* may be no more prominent than the transition in *frenzy* and *bronze* but will in this case be counted as a fully fledged consonant.

Realizing that speech and writing are different kinds or modes of expression, linguists have tried to establish phonetic segments from within the data of speech itself. This has not proved easy. Normal articulation does not consist of a series of discrete actions, each neatly separated from its neighbours. Acoustic records of speech as energy, obtained via a microphone, and articulatory records of the kind obtained by ciné X-ray film, electromagnetic articulography (EMA or EMMA) or ultrasound do not reveal tidily demarcated segments of speech; rather, they show a more or less continuous flow with various peaks and troughs of energy or movement. When listeners auditorily process this acoustic signal as speech, they are capable of interpreting it as a string of individually identifiable segments. Yet it is by no means obvious from an examination of this signal (and the articulatory activity which produces it) how the flow can be readily converted into a series of separate segments. We shall keep returning to this puzzle throughout this chapter, for the status of the segment as a unit has been questioned by some theorists. In the remainder of this section we shall be concerned particularly with ways in which phoneticians have tried to justify segmentation.

Even in investigations of articulation and perception it is impossible to ignore the units of speech. Of course, one could simply make measurements (of airflow rates or tongue movement or sound intensity) without any reference to units, but that would hardly constitute phonetics. In practice, phoneticians want to measure such things as tongue position in the production of certain kinds of vowels or consonants, or pressure variations during ejective stops. In so doing, phoneticians presuppose not only that there are indeed vowels and consonants of various kinds, but also that it is possible to tell where each segment begins and ends in the chain of speech. A simple working assumption is that, despite the continuous nature of speech, any sound can be identified as a stable state of the articulatory mechanism. This stable state is assumed to include all the articulatory settings that best characterize the sound in question, and is referred to by phoneticians as a TARGET.

Now some sounds, such as vowels, approximants and fricatives, can be fairly readily pronounced in isolation and prolonged indefinitely, subject to available air supply. For these sounds it does make sense to speak of a genuinely stable target – at least potentially, for the stable target will not necessarily be observable in running speech. Sounds such as stops, flaps, taps and trills, on the other hand, are inherently dynamic or transient in their articulation (section 2.12 above), and can be identified only by notional targets that relate

to characteristic articulatory properties. Despite this limitation, the concept of target remains important in phonetics and it is commonly used to justify canonical segments. Thus [k], for example, is said to represent a voiceless velar plosive without aspiration and with neutral lip position. This canonical definition may be convenient in an inventory of phonetic symbols and may also serve as a useful point of reference in studies of how velar plosives are actually articulated by speakers. It is none the less an idealization abstracted out of running speech, where the features of [k] – including even its voicelessness and point of articulation – may well vary considerably. However handy the concept of target may be, it remains important *not* to think of speech as a series of static targets linked by simple articulatory movements. If it has to be said that targets are rarely fully realized or are substantially modified in normal speech, then we must admit that the concept of target points beyond itself to assumptions about the organization of speech.

Besides making working assumptions, phoneticians have also tried to build analyses on the observable properties of the flow of speech. Since this flow does reveal some peaks of activity or prominence, it is theoretically possible to define units such as syllables and segments in terms of these peaks. A common approach proposes that a major peak of prominence represents the NUCLEUS of a syllable and that this nucleus will usually be a vowel or vowel-like segment; consonants will generally occur as MARGINS to these peaks, either as ONSET or as CODA. Other terms used in this kind of approach include CREST, with adjacent SLOPES (Jakobson and Halle 1956, p. 21), or PEAK, with adjacent TROUGHS (Jones 1960, pp. 55ff.). Authors using these terms concede that it is often difficult to locate the boundaries of segments and syllables on this basis; but, in a metaphor of Jespersen's (1922), one would not necessarily deny the existence of two adjacent hills simply because one cannot determine how much of the intervening valley belongs to either of them.

In general, there is no one physical variable which points unambiguously to segmental structure. Thus Stetson's contention that every syllable is initiated by a pulse of chest muscle activity – and that a syllable could therefore be defined by this criterion – was not substantiated by investigation of the muscles themselves (Stetson 1951, Ladefoged 1967).

Peaks of acoustic energy or of articulatory constriction are important in speech organization but again do not offer absolute criteria. As a rule, a peak of energy is likely to be identified as the nucleus of a syllable, and a trough of energy as the margin; while a peak (or trough) of constriction or stricture in the supraglottal tract is likely to be identified as the centre of a segment. In words such as *tot*, *tip-top*, *potato*, acoustic energy will be high at each vowel and sharply reduced during consonant articulation, supporting the impression of clear syllabic structure. At the same time, the degree of articulatory constriction which creates this variation in energy will obviously be highest during the consonants (which are all plosives, involving complete stoppage) and lowest during each vowel. Each segment can therefore be associated with a peak or trough of constriction (Pike 1943, pp. 107ff.). Nevertheless, there are instances where these criteria fail to produce the results expected by a native speaker. A word such as *gorilla*, for example, does not display the clear peaks

and troughs that can be observed in *potato*. Apart from its initial consonant, the word *gorilla* contains only vocalic and vowel-like (approximant) sounds. It shows a relatively even acoustic output and little variation in the degree of constriction. More seriously still, a word like *extra*, with a cluster of medial consonants pronounced as [kstr], displays reduced constriction between the [k] and the [t]. It might be expected that the peak of energy at this point, relative to the adjacent plosives, would count as a syllabic nucleus. But it is clear that English speakers count *extra* as a two-syllable word, not three; within English, the peak of energy at this point *does* count as a segment [s] but *not* as a syllable peak.

Another strategy of description resorts to SONORITY. The term refers to energy relative to effort, or more informally to the 'carrying power' of a sound. A sonorous sound is one with high output relative to the articulatory effort required to produce it, and sounds can therefore be ranked according to their degree of sonority. The vowel in *hawk* is more sonorous than the vowel in *hook*, which is in turn more sonorous than a consonant such as *l* or *r*. We can say in general that the points of greatest sonority in an utterance will be interpreted as syllable peaks. But the concept is again not entirely satisfactory. To the extent that it remains a measurable property, derivable from the speech signal, it falls foul of the same difficulties raised by the medial peak in a word such as *extra*; to the extent that it is redefined in more impressionistic terms it merely raises the question of the criteria that language users actually employ in judging segments and syllables.

It must be clear at least that language users who perceive an utterance to have a certain number of syllables or segments are doing more than processing physical data in a purely mechanical fashion, and the customary use of the term PROMINENCE indicates that various factors are integrated within a linguistic system. Thus sonorous sounds can be made less prominent – and non-sonorous sounds more prominent – by variations in duration, pitch or loudness, for example. Moreover, native speakers of a language operate within the system of that language – so that, for example, an English speaker's perception of *extra* never takes the medial [s] to be sufficiently 'prominent' to count as a syllabic peak. For reasons such as these, some phoneticians have abandoned the notion of the syllable altogether (Heffner 1964, Kohler 1966).

It must indeed be recognized that segmentation cannot be based on the speech signal alone but must be responsive to the systematic organization of the language as a whole. Our ultimate assurance that the English word *extra* has only two syllables or that *gorilla* has three is the fact that native speakers perceive these words in just that way. Although much of this complex and abstract process of native speaker perception remains poorly understood, we are not thereby obliged to resort to impressionistic speculation about syllables and segments. English speakers demonstrate the reality of syllabic structure by the way they count the syllables in a line of verse, for example; and they show that they can respond to segmental composition by their assessment of rhyme and alliteration, which depend on shared segments or groups of segments (*mat*, *bat*, *fat*, etc. or *mat*, *mop*, *mud*, etc.). Ultimately it is this kind of analytical insight – rather than mere inspection of physical records of peaks of energy and stricture – that justifies breaking such words into three segments.

It does not follow from this that all utterances in all languages can be unambiguously segmented, or that the location of boundaries between units is easily agreed. Given the continuous nature of speech and the overlap between successive articulatory events, it is usually easier to say how many segments or syllables an utterance has than to determine where exactly each unit begins and ends. Even in a simple case such as *door*, the transition from the [d] to the following vowel requires an articulatory movement of the tongue away from the tooth ridge, a movement which is simultaneously the end of the [d] and the beginning of the vowel. Comparable difficulties arise in determining syllable boundaries where, for example, medial consonants in words such as *falling* and *sugar* may be considered to belong both to the preceding syllable and to the following. Language users may or may not impose a sharp boundary, and may locate the boundary differently depending on the criteria to which they give priority. Thus when articulating the word *falling* slowly, as in shouting or singing, speakers may divide it as *fa-lling*, but when considering the grammatical structure, they may break it into *fall* + *ing*.

Our discussion so far has been almost entirely in terms of segments and syllables, reflecting the attention paid to these particular units in the literature, but other units are of importance in analysing speech. Larger units may be associated with patterns of stress, rhythm or intonation. The term PHONOLOGICAL WORD may apply not only to elements that also happen to be words in a grammatical or orthographical sense but more generally to closely knit clusters of syllables such as *an apple*, *at home*, *doesn't* or *shouldn't have*. Terms such as BREATH GROUP and PAUSE GROUP may refer to the 'chunks' into which we divide speech by pausing or taking a new breath, and INTONATIONAL PHRASE commonly refers to a unit defined within the intonational system (see further chapter 9 below).

In summary, segmentation and structural organization are crucial notions in phonological analysis but are not recoverable from orthographic practice or instrumental records of speech alone. Peaks (or troughs) of energy or stricture *tend* to be identified as the central points of units such as segments and syllables but are not absolute criteria in phonological analysis. The determination of phonological units – and in particular the location of boundaries between units – requires reference to the phonological system as a whole.

3.2 Complex articulations

Segmentation of the speech chain and identification of individual sounds are complicated by the way in which many languages use sounds involving combinations of articulatory values. Thus, for example, while oral airflow is the normal prerequisite for a vowel, certain vowels in some languages may have nasal airflow as well (section 2.7). These vowels are normally termed NASALIZED, suggesting that the nasalization is a secondary or superimposed articulation. There are also sounds which use two places of articulation simultaneously, a

common example being the [w] heard in English *win* or *wet*, an approximant with narrowing at both lips and velum.

Phoneticians commonly make a distinction between so-called SECONDARY ARTICULATIONS (such as vowel nasalization) and other forms of complex articulation involving more than one place of articulatory activity in the vocal tract. Pike (1943), for example, defines secondary articulations as those associated with constrictions ranking lower than main articulations. But it is sometimes a matter of debate whether one articulator is subsidiary to another. Recognizing this, Ladefoged (1971) suggested that secondary articulation could include other complex articulations if we arbitrarily assigned primary ranking to the stricture closest to the glottis.

Neither Pike nor Ladefoged deals satisfactorily with the dimension of time. Secondary articulations and other complex articulations may both show either simultaneous components or transitional modifications, and this further blurs the distinction between them. As there seems to be no adequate basis for a rigorous distinction between secondary and complex articulations, they are treated here as a single category under the heading of COMPLEX ARTICULATIONS, which is the more general and potentially less misleading term. We recognize two types of complex articulation: SIMULTANEOUS (separate but co-occurring articulatory activities which result in the production of a sound identifiable as a single segment) and TRANSITIONAL (separate and successive articulatory activities which together can be identified as a single segment).

There are special symbols for some complex articulations but in most instances either a digraph (composed of two appropriate symbols) or a diacritic is used. The digraph may if necessary be distinguished from a sequence of two segments by the use of a linking line above it, e.g. [t͡ʃ] in English *chop* or *much* versus [tʃ] in *hat-shop* or *hot-shot*. Diacritics for simultaneous complex articulations are placed on, above or below the main symbol, as in [ã] for a nasalized [a]. Transitional complex articulations may use either a superscript or a digraph, where the choice can reflect the relative prominence of one or other aspect of the complex articulation, as in [tˢ] or [t͡s].

3.3 Nasalization

We have already noted that vowels can be nasalized by lowering the velum to allow air to flow through the nasal cavity as well as through the oral cavity (section 2.7 above). This is SIMULTANEOUS NASALIZATION. It is also possible to nasalize consonants such as approximants and fricatives, but this seems to happen only as a response to context: for instance, a [w] standing between two nasalized vowels in the west African language Yoruba will normally be nasalized because the nasalization is continued through from one vowel to the next. Nasalization may often spread across several segments in comparable non-significant ways: in English, for example, a word such as *channel* may be pronounced with nasalization of both the vowel preceding [n] and the lateral following [n]. We indicate nasalization with the diacritic known as a tilde, thus: [ã], [ũ], [w̃].

Nasalization may be described as 'inherent' when speakers do not exert strong control over the raising of the velum, allowing nasalization to become an 'unintended' characteristic of all their vowels, even when not adjacent to nasal consonants. Nasalization may also be a general property of speech, for reasons of individual articulatory habit, dialect type, or pathological condition such as a cleft palate. Such nasalization is often described as 'pervasive'.

In PRENASALIZATION, a component of nasal articulation occurs before, or in the initial part of, the basic articulation of a segment. Most commonly this applies to stops, which can have nasal output during the initial part of the occlusion phase. Fijian *b* and *d*, for example, are prenasalized and may be represented in phonetic transcription as [m͡b] or [ᵐb] and [n͡d] or [ⁿd].

POSTNASALIZATION is the sequential reverse of prenasalization, with transitional nasal coupling at the end of the basic articulation. The central Australian language Aranda, for example, has stops that may be represented as [p͡m] or [pᵐ], [t͡n] or [tⁿ], and so on.

Note that the standard terminology is somewhat confusing. Nasal consonants (section 2.12 above) might be thought of as nasalized stops, since they have nasal airflow accompanying the articulation of a stop; but it is not usual to describe them in this way and the reader should be aware of the customary distinction between nasal consonants (i.e. nasal stops such as [m] and [n]) and nasalized sounds (especially vowels but also fricatives and approximants). Nevertheless, in stops which are traditionally described as 'prenasalized' and 'postnasalized', the so-called nasalization is in fact a brief nasal stop.

3.4 Labialization

Labialization is the addition of lip rounding or lip protrusion to any sound which is normally articulated with the lips in a neutral or spread position. Labialization modifies the basic articulation by extending the length of the vocal tract and altering its cross-section.

In SIMULTANEOUS LABIALIZATION, rounding or protrusion occurs during the basic articulation. In English (as in most languages) the lip setting of vowels is likely to be maintained through adjacent consonants, particularly where a consonant stands between two identical vowels. Thus the name *Lulu*, for example, may be pronounced with lip rounding persisting through the second [l] sound. This simultaneous labialization is marked by a subscript [w], as in [l̪].

In TRANSITIONAL LABIALIZATION, the rounding or protrusion is most evident at the end of the main articulation as a part of the transition to the next segment. Not uncommonly, velar stops may have distinctive labialization of this kind, often shown by spellings such as *kw* or *qu*. The phonetic representation is with a superscript, e.g. [kʷ].

3.5 Palatalization

Palatalization involves raising the tip and blade of the tongue to a high front position close to the anterior part of the hard palate region, as for an [i] vowel.

In SIMULTANEOUS PALATALIZATION, the modification to tongue position occurs at the same time as the other articulatory gestures of the segment. An alveolar lateral, for example, can be produced with the body of the tongue (behind the lateral stoppage) raised towards the [i] vowel position. This yields what is often described as a 'clear l', as opposed to the 'dark l' with the body of the tongue lower and further back, as for an [u] vowel. The difference may be heard in comparing the standard German and English pronunciations of names such as *Helmut* and *Wilhelm*, in which the [l] sounds will normally be clear in German but dark in English.

Other sounds that can be palatalized in this way include [n] and [r]; some sounds, such as [k] and [g], cannot have simultaneous palatalization because of the requisite tongue position. We represent simultaneous palatalization by a subscript [j], as in [l_j].

In TRANSITIONAL PALATALIZATION, the constriction of the basic articulation is released through a palatal approximation of the tongue tip and blade, as part of the transition to the next segment. In the articulation of stops, the approximation may sometimes be so close that it causes a degree of airstream turbulence (and hence affrication) in the release. It is probably more common than simultaneous palatalization and occurs with a wide variety of consonants in Slavonic languages such as Russian and Polish. The phonetic representation of, for example, Russian or Polish palatalized [t] is [t^j].

3.6 Velarization and pharyngealization

Velarization and pharyngealization involve moving the tongue body and root from their neutral vocal tract position towards the positions for the vowels [u] and [ɒ]. Since the tongue body posture is adjusted, these articulations always occur simultaneously with the basic articulatory gesture. The so-called 'dark l' referred to above in contrast with (simultaneously palatalized) 'clear l' is velarized to some extent. In Arabic, the so-called 'emphatic' consonants are either velarized or pharyngealized. There appear to be no languages which employ both velarization and pharyngealization. A common practice, followed here, is to place a tilde through the main symbol as a general marker for either of the two complex articulations, e.g. [ɬ].

3.7 Affrication

There is almost always some degree of air turbulence (and hence friction) at the release of a stop. This is normally of such short duration that it counts as part of the release burst of the stop itself. But when the release is strongly frictional and is extended in duration, it can be identified as a separate fricative phase of the articulation. A single complex segment of this kind, in which the articulators release an occlusion through a controlled fricative phase, is known as an AFFRICATE or AFFRICATIVE. The constriction of the fricative phase may be central, grooved or lateral (section 2.13 above).

We can distinguish two kinds of affricate, according to the strength and duration of the frictional release. We use the term AFFRICATED STOP for the weaker kind, and reserve the term AFFRICATE for the stronger version. In London English, the [t] of e.g. *ten* or *time* may be heavily aspirated to the point of affrication (Gimson 1980, p. 160). The phonetic representation of this affricated stop is [t^s]. This can be distinguished from a true AFFRICATE, such as the [t͡s] of German *zehn* or *zwei*.

3.8 Double articulation

Two stop articulations can be made simultaneously, creating a double stop. The most widely quoted instances are the labial and velar stops (sometimes called labio-velars) of west African languages such as Yoruba and Ewe, which can be represented as [k͡p] and [g͡b]. Note that these are true double stops, with simultaneous occlusion, not sequences as in English *crackpot* or *ragbag*. Other places of articulation can be combined, and while Ladefoged (1971) suggests that one of the two closures must always be labial, there seems to be no physiological basis for such a restriction. We include as double stops various stops with simultaneous glottal closure, such as [t͡ʔ] or [k͡ʔ]. In some varieties of English, speakers tend to produce syllable-final voiceless plosives (as in *sit* or *sick*) in this way. The replacement of the final [t] by a glottal stop (characteristic of some Londoners' speech, for example) can be regarded as an extreme development of this tendency to use glottal closure in stops: the tongue gesture of the double stop [t͡ʔ] disappears entirely. The mechanism of a double constriction is not limited to plosives and may apply also to nasals, e.g. [ŋ͡m] (the nasal analog of [k͡p] and [g͡b]).

3.9 Vowel retroflexion

In vowel retroflexion, the basic tongue posture of the vowel is modified to produce an 'r-colouring' of the auditory quality. Many American and Irish speakers will have 'r-coloured' vowels in words such as *far*, *four* and *fir*. The auditory effect is traditionally thought to result from curling the tip and blade of the tongue back from their normal vowel position without touching the roof of the mouth, while retaining the basic tongue root and body position for the vowel concerned. These 'r-coloured' vowels are sometimes described as RHOTACIZED vowels.

Ladefoged (2006, p. 92) suggests from X-ray evidence that there may be another way of producing rhotacization, by leaving the tongue tip down but bunching the tongue body upwards. He claims that both articulatory strategies cause the root of the tongue to retract, narrowing the pharynx. Vowel retroflexion is normally regarded as a simultaneous modification of the basic vowel articulation, but in many instances it occurs only in the coda of a vowel, and may thus be transitional in nature. We may represent all varieties of rhotacization by the same superscript symbol, e.g. [a], [ɔ˞], but where the transitional effect is very prominent, it may be more appropriate to interpret the articulation as an approximant consonant following the vowel, i.e. [aɹ], [ɔɹ].

3.10 Diphthongization

We have already referred to diphthongization as a complex articulation combining pure vowels (section 2.8 above). The timing of articulatory movement between two endpoints of pure vowel quality can be finely varied, but three categories appear adequate to account for auditory distinctions. ONGLIDE refers to a relatively brief onset leading into a dominant vowel quality, as in, for example, the [ᵊi] heard from many Londoners and Australians in the words *sea* and *me*. OFFGLIDE refers to a comparable effect at the end of a vowel, moving away from the dominant vowel quality. Some speakers of conservative RP and some from the southern USA have [ɔᵊ] in words such as *pore* and *lore* (as opposed to *paw* and *law* without the offglide). The term DIPHTHONG is then reserved for a glide between two vowel qualities, neither of which dominates, e.g. [ɔɪ] as in *toy* or *coy* (in virtually every variety of English).

This categorization of glides and diphthongs is based on the major auditory distinctions that divide the continuum of articulatory timing. In the following sections, we shall see that diphthongs – in common with most of the complex segments mentioned in this chapter so far – need further assessment in the light of the functional role which they play in particular linguistic systems.

3.11 Syllabicity

A syllable commonly consists of a vocalic peak, which may be accompanied by a consonantal onset or coda. In some languages, *every* syllabic peak is indeed a vowel. But other sounds can also form the nucleus of a syllable. In English, this generally happens where a word ends in an unstressed syllable containing a nasal or lateral consonant, as in the following examples, where syllabic consonants are marked by the conventional subscript:

sudden [sʌdən] or [sʌdn̩]
medal [mɛdəl] or [mɛdl̩]

To take another example from English, the word *and* may often be pronounced as a syllabic nasal (as reflected in the informal spelling *'n*). In fast colloquial speech, the nasal consonant may be assimilated to [m] before a bilabial consonant or [ŋ] before a velar, although this is often condemned as slovenly or careless:

bread 'n butter: 'n may be [n̩] or [m̩]
cash 'n carry: 'n may be [n̩] or [ŋ̩]

In German, unstressed final *-en* is frequently pronounced as a syllabic nasal. Here, the nasal may assimilate to the point of articulation of the preceding consonant, although again the pronunciation is often condemned (e.g. Siebs 1961, p. 43):

haben	('to have')	[ha:bən] or [ha:bm̩]
geben	('to give')	[ge:bən] or [ge:bm̩]
sagen	('to say')	[za:gən] or [za:gŋ̩]
danken	('to thank')	[daŋkən] or [daŋkŋ̩]

Syllabic consonants can be identified as such by their relative length and in some instances by the lack of any audible vocalic release of the preceding stop. (Compare the [d] in the two pronunciations of *sudden*.) In the majority of cases in English, there is variation between syllabic and non-syllabic forms of these consonants. Where they occur medially, there may even be variation in the number of syllables in the word: *fiddler*, for example, can be three syllables (*fidd-l-er*) or two (*fidd-ler*). Ladefoged (1971) suggests, however, that the difference may sometimes be distinctive. His example is *coddling* (three syllables, from 'coddle') versus *codling* (two syllables, meaning 'baby cod').

Syllabic nasals are reasonably common in African languages. In Swahili, for instance, the word for 'man' is *mtu* [m̩tu]; in Yoruba, 'big' is *ńlá* [ń̩lá]. Both words are pronounced as two syllables. (The accents in Yoruba mark tone, which occurs on the syllabic nasal as well as on vowels.)

Less commonly, other consonants besides nasals and laterals may have a syllabic value. Syllabic fricatives are reported from languages of the Pacific Coast of North America, such as Bella Coola, a language of British Columbia in which the word for 'bad' is [ʃχ] and 'northeast wind' is [sps̩] (Hoard 1978, p. 67).

Syllabicity is sometimes treated as an additional articulatory quality – so that nasal consonants, for example, are specified as either nonsyllabic or syllabic. While this may be convenient in description, it should be noted that perception of syllabicity is shaped by the perceiver's linguistic system. English speakers take the word *spots* to be a single syllable; Bella Coola speakers, on the other hand, might well take it to be a three-syllable word, since it has three phonetic crests, with the initial and final fricatives forming peaks relative to the adjacent plosives.

In the case of some sounds, syllabic and nonsyllabic counterparts tend to be identified as vowels and consonants. This is particularly true of the vowels [i] and [u] and the corresponding approximants [j] and [w]. The term SEMIVOWEL, used for approximants such as [j] and [w], points to the fact that the distinction between a syllabic high vowel and a nonsyllabic semivowel is not always clear-cut. Orthographic practice is often confusing. Classical Latin orthography, for example, used *i* to represent both [i] and [j] and *u* to represent both [u] and [w]. Although the Romans themselves considered the possibility of introducing a distinction in spelling, it was not until the Renaissance that the custom of using *i* and *u* for the vowels and *j* and *v* for the consonants was introduced into the writing of Latin, e.g. *major* 'greater' for earlier *maior*, *pluvium* 'rain' for earlier *pluuium*, etc. (The variants *i/j* and *u/v* were originally different styles of the same letters; see Allen 1978, p. 37.) In English, there are alternative pronunciations of words such as *piano* and *fiasco*, depending on whether they are pronounced with the vowel [i] in the initial syllable (making them three-syllable words) or with the semivowel [j] following the initial consonant (in which case they have only two syllables).

Other semivowels in addition to the widespread [w] and [j] include a labio-palatal [ɥ], which is the approximant counterpart of the front rounded vowel [y], and a velar [ɰ], which is effectively an unrounded [w], corresponding to the back unrounded vowel [ɯ]. French has both [w] – as in *oui* [wi] 'yes' or *loi* [lwa] 'law' – and [ɥ] – as in *nuit* [nɥi] 'night' or *puis* [pɥi] 'then'. The velar [ɰ] is rather rare but occurs in Karen (Burma) and in some Australian Aboriginal languages (where it sometimes seems to have been derived from [w] by a process of derounding or from [ɣ] by loss of friction).

Certain kinds of approximant *r* can be also viewed as semivowels corresponding to retroflexed vowels (section 3.9 above). As in other instances mentioned above, there is often doubt as to the true status or value of these segments. In some varieties of American English, for example, sequences of *r* + vowel and sequences of vowel + *r* merge into a single syllabic sound which could be regarded as either a retroflexed vowel or a syllabic approximant [ɹ]. Thus the words

pretty prevent purpose pervert

may all be pronounced with an identical initial syllable [pɹ̩-]. Note that spellings such as *purty* or *prevert* represent the typical spelling mistakes of a speaker who makes no distinction in pronunciation between *pre-* and *per-* (although these 'misspellings' are often deliberate and jocular nowadays, rather than genuine errors).

3.12 Segmentation and structure

The segmental organization of some languages can be relatively easily described. In Walmatjari, for example, an Aboriginal language spoken in the north of Western Australia (Hudson 1978, pp. 4ff.), every word begins with a single consonant; consonants may be adjacent to each other only at the junction of two syllables; and the boundaries of segments and syllables are fairly easily determined. Thus for each of the following words we can show the separate syllables and their structure in terms of C (consonant) and V (vowel):

ngapa ('water')	[ŋa pa]	CV CV
kurrapa ('hand')	[ku ra pa]	CV CV CV
ngarpu ('father')	[ŋaɹ pu]	CVC CV.

We stress that the description is *relatively* easy and that there are some problems (which will be mentioned later). But it is generally possible to identify units such as segments and syllables and to give a relatively straightforward account of their patterning. Thus we can say that a Walmatjari syllable is realized as either CV or CVC, that words must end with a CV syllable, and so on.

The term DISTRIBUTION is often used in such contexts and a description of phonological structure is sometimes known as a DISTRIBUTIONAL STATEMENT (i.e. a statement specifying how segments are distributed within syllables, and syllables within words, etc.). The term PHONOTACTICS is also widely used to refer to the general description of sequences and combinations. In particular, PHONOTACTIC CONSTRAINTS express limitations on the free combination of units. In English, for example, a word-initial syllable can begin with various consonant sequences (such as *sp*, *st*, *sk*, *pl*, *bl*, *kl*, *gl*) but there are systematic constraints that exclude *zb*, *zd*, *tl* or *dl*. Sequences of three consonants are even more severely constrained in English: in initial position we have, for instance,

spr as in spring, sprat
str as in string, strap
skr as in script, scrap

but not *smr*, *zbr*, *fpr*, etc. (See Gimson 1980, pp. 237 ff., for a detailed account.)

Accounts of distribution or phonotactics sometimes suggest an unduly mechanical picture, as if segments were building blocks put together to form

syllables, and syllables in turn put together to form larger units. In fact phonotactic description cannot be divorced from the process of segmenting speech and establishing units, and there are often difficulties even in distinguishing between a single segment and a sequence of segments. This is particularly true of complex articulations such as prenasalized plosives or affricates (sections 3.3 and 3.7 above), where we seem to be talking about single segments even though the articulation indicates a sequence of two elements or components.

In some cases a language may actually distinguish between one-segment and two-segment articulation. Most English speakers will probably regard the initial consonant of *chip*, *chain*, *chop* as a single consonant, even though it is an affricate consisting of a stop and a fricative. Certainly there is no possibility in English of distinguishing between a word-initial affricate and a word-initial sequence of stop + fricative. But in cases such as *he cheats* versus *heat-sheets* or *what can each add?* versus *what can eat shad?* there is, at least potentially, a distinction (Chao 1934, p. 40). Readers may like to consider whether and how they distinguish between the medial consonants in such pairs of words as

lychee	light-ship
ketchup	pet-shop
urchin	(a) hurt shin

A direct contrast among such words in connected speech is of course unlikely, but many speakers of English will sense a difference, principally in the timing of the transition from the stop to the following fricative.

More commonly, there is no distinction within the language: some languages have affricates, others may have sequences of plosive plus fricative. In such cases what appears to be identical in articulation proves to be differently valued in different languages. The sequence [ts], for instance, occurs in English – *cats*, *fatso*, *Betsy* – Dutch – *fiets* ('bicycle'), *etsen* ('to etch') – and German – written *z* or *tz* in *zehn* ('ten'), *bezahlen* ('to pay'), *salzen* ('to salt'), *witzig* ('funny'). In English and Dutch this sequence is not an affricate but simply a sequence of [t] and [s]; in German, however, [ts] is normally counted as a single-segment affricate, just as [tʃ] is in English.

Affricates of this general type (plosive plus sibilant fricative) are reasonably common and are often represented as single letters within the spelling system of the relevant language, e.g.

[ts]	Polish and Czech *c* in *co* (what) *cena* (price)
	Italian *z* in *zio* (uncle) *marzo* (March)
[dz]	Italian *z* in *zona* (zone) *zero* (zero).

Russian script also has a single letter for [ts], while in ancient Greek the letter zeta seems to have represented [dz]. (In modern Greek, zeta represents [z], but it is thought that the Ancient pronunciation was [dz], metathesized to [zd] early in the Classical period; see Allen 1987, pp. 56–9.)

Affricates similar to English [tʃ] (*chip*, *chess*) and [ʤ] (*jump*, *jeep*) are also widespread, but with variation in the precise point of articulation. In the Pinyin romanized spelling of Chinese, for example, *q* and *j* represent palatal affricates, and *ch* and *zh* postalveolar affricates, as in *qiang* ('rob'), *jian* ('build'), *chuanghu* ('window'), *zhui* ('chase'). In Indonesian, *c* and *j* represent sounds that are sometimes described as palatal plosives but are normally articulated with a clear fricative release, as in *cari* ('search'), *jari* ('finger') or names such as *Cilacap*, *Jawa* (Java), *Jakarta*.

Other less common types of affricate include [pf] as in German *Pferd* ('horse') or *pflanzen* ('to plant') and [kx], a velar affricate which occurs in some Swiss German dialects where the standard language has initial *k* (Russ 1978, p. 46). An affricate with lateral fricative release is found in a number of indigenous American languages, represented as *tl* in names such as *Nahuatl* and *Tenochtitlan*. In all such instances, the definition of these complex sounds as affricates depends upon the single-segment value assigned to them within the particular phonological system (section 3.14 below).

Similar remarks can be made about prenasalized stops. Many languages allow sequences of nasal plus plosive (English *amber*, *under*, *anger*) but it is only in certain languages that such sequences have the status of single segments. A clear instance is Fijian, where [mb], [nd] and [ŋg] count as single consonants within the phonological system. Voiced plosives are always prenasalized in Fijian and [b], [d] and [g] do not occur other than in prenasalized form. Note, for example, the Fijian spellings *Nadi* (place name) and *noda* 'our' for [nandi] and [nonda].

Lengthened sounds are also open to interpretation as single or double sounds. In English, there is no particular reason to take long vowels to be sequences: some vowels are relatively long (as in *beat* and *boot*) and others are short (as in *bit* or *but*). Depending on the variety of English, some long and short vowels can be paired as vowels distinguished by length alone, but others cannot. In Australian English, for example, the long vowel of *calm* and *psalm* can be matched with the short vowel of *come* and *sum*: the two vowels are distinguished by length alone. But the long vowel of *bead* and *fees* has no short partner. (If there is a long counterpart of the short vowel of *bid* and *fizz*, in typical Australian pronunciation it is the long vowel of *beard* and *fears* rather than that of *bead* and *fees*.) On the other hand, in Japanese, there are five short vowels and five matching long ones. Not only do the long vowels count as double vowels in written Japanese, they are also reckoned as two vowels in the rhythmic organization of verse (Comrie 1987, p. 868).

3.13 Diphthongs and related phenomena

At first sight diphthongs seem to be another instance of two phonetic segments functioning as one. They are, in a sense, two vowels forming a single entity (sections 2.8 and 3.10 above) and are analogous to affricates or prenasalized plosives in that they are generally regarded as single but complex segments. Nevertheless, the interaction of syllabicity and segmentation can be intricate.

We may begin by distinguishing two simple categories, namely diphthongs and sequences of two vowels. English vowels heard in the RP pronunciation of the following words are undoubtedly diphthongs:

how, bough, cow
eye, high, buy, tie
hay, bay, cay
owe, hoe, toe
boy, toy, coy.

All of these words count as single syllables in English. On the other hand, there are languages in which sequences of vowels are clearly not diphthongs. In Komering, for example, a language of southern Sumatra, words such as *mait* ('corpse'), *tuot* ('knee') and *kuah* ('sauce') are articulated and perceived as two-syllable words (Yallop and Abdurrahman 1979, pp. 11–12). Note that syllabicity is crucial in distinguishing the English diphthongs from the Komering two-vowel sequences.

Now in the case of diphthongs, the first or second target is often a vowel of [i] or [u] quality. In some languages, especially where the [i] or [u] target is clearly not the dominant component of the diphthong, the possibility arises of treating the [i] or [u] as a nonsyllabic semivowel. Thus diphthongs such as [oi] and [au] might be represented as [oj] and [aw]. In fact notations of just this kind have been widely used in English transcription, particularly in the USA, e.g. *high* as [haj] and *how* as [haw].

With sequences of two vowels, transcription is complicated by the fact that an intrusive glide may be heard between the two vowels. Thus in many languages, words such as [mei] or [mou], articulated as two syllables, may well be heard as [meji] or [mowu].

It may be necessary to make a further distinction here, as there are also 'triphthongs' in some languages, vowel sequences in which three components can be heard but which none the less count as a single vowel. In some varieties of English (notably those without postvocalic *r*) words such as *hire*, *lyre* and *our*, *cower* contain triphthongs, often transcribed as [aɪə] and [aʊə]. Here too, however, intrusive glides may occur, and some speakers may distinguish between, e.g.

one syllable:	hire	lyre	flour	cowered
two syllables:	higher	liar	flower	coward.

But many authorities, including Jones and Gimson, do *not* recognize any distinction between these pairs, treating every one of these eight words as a monosyllable containing a triphthong.

It is also possible to analyse long vowels as vowel + semivowel. Hence [i:] or [u:] may be transcribed as [ij] or [uw], particularly if there is a change in quality during the articulation of the vowel, or if a transitional glide becomes prominent before a following vowel, that is, if [i:] followed by [a] is pronounced as [ija]. Note, for example, the English tendency to insert a [j] glide after the

first vowel of words such as *piano*, *Fiona* and *Seattle*. Again, there is a tradition, notably in the USA, of transcribing English long vowels in this way.

In some Aboriginal languages, phonological structure suggests an interpretation which is the mirror-image of the above. That is, [i:] may be interpreted as [ji]. In Walmatjari, for example, we saw (in section 3.12 above) that words regularly begin with a consonant. There are apparent exceptions to this pattern, specifically words beginning with [i:] and [u:], such as [i:nja] 'gave' and [u:lju] 'good'. But these words are not exceptions if they are taken to begin with [ji] and [wu]. If so, then [i:nja] is [jinja] and [u:lju] is [wulju]. While this may seem a surprising equation, it is important to see how such interpretations make sense within a particular system. Unlike English, Walmatjari has no contrast of the type found in English *east/yeast*, *ooze/woos* and no words beginning with a vowel (other than the 're-interpretable' instances of initial [i:] and [u:]). Hence the predominant pattern of Walmatjari – that words regularly begin with a consonant – may be sufficient to make speakers of the language 'feel' that [i:nja] and [u:lju] do also begin with a consonant.

We can therefore distinguish, potentially at least, a number of distinct structural possibilities:

1 DIPHTHONG, i.e. a single vowel, but one in which two targets or components can be discerned. The term is warranted only if the vowel genuinely counts as a single vowel in the language in question. In English, for example, *high*, *how* and *hoe* are judged by speakers to be monosyllables and their vowels can justifiably be called diphthongs.
2 TRIPHTHONG, i.e. a single vowel with three discernible targets. As with diphthongs, the term presupposes some justification for treating the complex as a single vowel.
3 VOWEL PLUS SEMIVOWEL, e.g. [ej] [aj] [ow] [oj]. In many languages such sequences are simply alternative ways of transcribing diphthongs, i.e. [ej] = [ei]; but there are Australian Aboriginal languages in which the second component of a diphthong such as [ai] has a consonantal value, and is therefore analysed as [j]. Where a high vowel is followed by an appropriate semivowel (high front vowel followed by palatal semivowel, high back vowel followed by velar semivowel) it is unlikely that the sequence will be distinct from a long vowel. Hence [ij] = [i:] and [uw] = [u:].
4 SEMIVOWEL PLUS VOWEL, e.g. [ju] [wa] [ji] [wu]. As with vowel plus semivowel, these sequences may be alternative transcriptions of diphthongs or (where appropriate) long vowels.
5 VOWEL SEQUENCE, i.e. a sequence of two (or more) vowels which is *not* a diphthong. If such a sequence is genuinely distinct from the other possibilities listed above, it is highly likely that the constituent vowels will constitute separate syllabic peaks.
6 VOWEL PLUS SEMIVOWEL PLUS VOWEL, e.g. [aji] [uwa]. In some cases these will simply be part of the phonotactic possibilities of the language; in others the semivowel may be regarded as an intrusive transition. Compare English words like *leeway* and *blow-wave*, where the semivowel [w] readily counts as the initial consonant of the second syllable; and words like

booing, *cluey*, *blowy* and *Joey*, where a medial [w] may be taken to be intrusive.

The ways in which these six possibilities are exploited and equated with each other vary enormously from language to language, and it is important to avoid generalizations based on one or two languages. Thus the tendency to link adjacent vowels via a semivowel is strong in some varieties of English (as in *seeing* with intrusive [j], *suing* with intrusive [w]); and the consequence of this is to blur the distinction between [i:] and [ij]. In German, however, there is a quite different tendency, namely to separate adjacent vowels by a glottal stop: compare German *Hiatus*, pronounced [hia:tus] or [hiʔa:tus], with the English pronunciation of *hiatus*.

Such tendencies are by no means predictable from general phonetic principles and may reflect complex linguistic patterns. Thus some speakers of English extend the pattern of intrusive semivowels to include insertion of an r-sound between certain vowels – the so-called 'intrusive r' in e.g. *law(r) and order*, *the idea(r) of it*, *draw(r)ing*, etc. This r-transition applies only after certain vowels, namely [ɜ:] [ɔ:] [a:] [ə] and centring diphthongs such as [ɪə] and [ɛə], and is restricted to varieties of English in which *r* is not normally pronounced unless followed by a vowel. In such varieties (e.g. south-eastern England, Australia, New Zealand, north-eastern USA) we find:

r is not pronounced at the end of an utterance,
  e.g. two plus two equals fou(r);
r is likewise not pronounced before a consonant,
  e.g. fou(r) books; fou(r) tables;
but *r* is pronounced before a vowel,
  e.g. four apples; four eggs.

This pattern sets up a powerful analogy for *all* words ending in an appropriate vowel, including those in which there is no *r* in the spelling, such as *law* and *idea*. The 'intrusive r' is thus a consequence of the historical loss of word-final and preconsonantal *r* in certain regions of the English-speaking world. The 'intrusion' is quite unnatural for those who retain the *r*, including most Scottish and Irish and many North American speakers of English. Hence judgements about diphthongs, semivowels and transitions must in general take careful note of the particular phonological system within which the phenomena are found.

3.14 Interpretations

Questions of the kind that have arisen in this chapter – whether two consonants truly constitute an affricate or whether a vowel is really a semivowel – have generally been treated as questions of interpretation. The assumption here

is that certain sounds (or combinations) will need to be interpreted within the linguistic system of which they are part. The concept of interpretation has a long history. De Saussure, for example, proposed a phonetic classification that allowed certain sounds to function as either vowels or consonants (more strictly, in his own French terminology, as *sonantes* or *con-sonantes*). In his illustration from French, the *i* in both *pied* and *fidèle* is a single phonetic 'species' but [j] in *pied* is *consonante*, [i] in *fidèle* is *sonante* (1916, pp. 87–8). Some years later, Sapir suggested various criteria by which one might determine what he calls 'the place of a sound in a phonetic pattern over and above its classification on organic or acoustic grounds' (1925, p. 19). The criteria hinted at by Sapir and later exploited in the phonological description of many different languages include the following.

Combinatory possibilities or phonotactic patterning

For example: English [tʃ] and [ʤ] (as in *chin*, *chart*, *gin*, *jump*) are single-segment affricates because of their occurrence in word-initial position. In general, English words cannot begin with a combination of stop followed by fricative. (Note that where English spelling appears to allow word-initial stop plus fricative, the pronunciation is of a fricative alone, e.g. *ps* pronounced [s] in *psychology*, *pseudo*, *x* pronounced [z] in *xylophone*, *xenophobia*.) Now if English [tʃ] and [ʤ] were sequences of two consonants, they would violate this general pattern; but if they are taken to be affricates, the generalization that English words do not begin with stop plus fricative remains valid.

Patterns of stress or other prosodic regularities

For example: in Alyawarra, a central Australian language (Yallop 1977), words may contain sequences such as [pmp] or [tnt], e.g. *apmpima* ('burn'), *atntirrima* ('run'). The nasal in these sequences appears to form a syllabic peak, but the stress system overrides this impression and suggests a nonsyllabic interpretation. The general rule in Alyawarra is that words which begin with a vowel are stressed on the second syllable, e.g. *a'tirra* ('cicada'), *a'nima* ('sit') etc. Now in words such as *apmpima*, the stress falls not on the syllabic nasal but on the following vowel, i.e. *apm'pima*, *atn'tirrima*. Hence within the Alyawarra phonological system, the nasal consonants do not count as syllabic peaks and are better interpreted as release features of the preceding plosive. In other words, *pm* and *tn* are complex segments rather than sequences of two consonants.

Symmetry and parallelism

For example: Moba, a language spoken in Togo, west Africa, has a number of short and long vowels, illustrated in the following:

[bil]	to put	[bi:]	to be spoilt
[pel]	to hurry	[kud]	to prepare porridge
[tud]	to push	[tu:d]	to trip over
[pal]	to clean	[kod]	to slaughter
[pɔl]	to plug	[ka:d]	to interrogate the dead.

Now there is an asymmetry in the vowel system, for there are no long vowels matching short [e] [o] [ɔ]: we have

long vowels	i:	a:	u:
short vowels	i	e a o ɔ	u

There are, however, diphthongal vowels [ie], [uo] and [ua], as in

[piel]	to harvest peanuts	[miel]	nose
[puod]	to cross	[kuod]	to diminish
[pual]	to pluck	[kuad]	to sell.

If these diphthongs are interpreted as long vowels – [ie] = [e:], [uo] = [o:] and [ua] = [ɔ:] – the gaps in the system are filled. It must then be assumed that Moba does have long mid vowels (in a functional or systemic sense), but that these vowels are realized as diphthongs with a high onglide (Russell 1980).

Morphological or grammatical patterning

For example: in many Australian Aboriginal languages, there are often phonotactic reasons for interpreting certain vowels as semivowels. In some instances such interpretations are supported by the rules or patterns of affixation. In Dyirbal, for example, words can end in [ui], which is probably best interpreted as [uj], e.g. [walguj] 'brown snake' rather than [walgui]. An important criterion here is the affixation: the locative suffix in Dyirbal takes the form [ŋga] after a vowel but [ɟa] after a palatal consonant, e.g.

[jaɹa]	man	[jaɹaŋga]	on a man
[biɲɟiriɲ]	small lizard	[biɲɟiriɲɟa]	on a lizard

Now the locative form of 'brown snake' is [walguiɟa], not [walguiŋga]. Even though the root-final [i] may strike the hearer as clearly vocalic, it seems reasonable to say that, from the Dyirbal perspective, the root ends in a consonantal [j] (Dixon 1972, p. 42).

The most highly formalized statements of criteria such as these can be found in Trubetzkoy (1939, ch. 2) and Pike (1947, ch. 12). Trubetzkoy proposes a series of rules for what he calls 'monophonematic' interpretation (i.e. interpretation of complex articulations as single segments) and 'polyphonematic interpretation' (i.e. interpretation of a single segment as more than one). His two most important rules or conditions are as follows:

1 If two segments are to be interpreted as one, they must fall within the same syllable.
2 If two segments are to be interpreted as one, they must involve a unitary articulatory movement, e.g. [ts] and [kx] are potential affricates, but [ks] and [tf] are not.

Although these two rules have an air of common sense about them, they are not entirely satisfactory. A recurrent theme of this chapter has been that the definition of the syllable and the determination of segmental boundaries interact with each other, and it will not always be self-evident that two sounds fall within the same syllable. Furthermore, while there are obviously limits on the phonetic nature of potential diphthongs and affricates, these limits are not easily defined. We know, for instance, that [pf] and [tʃ] count as affricates in some languages, and we must presumably define 'unitary articulatory movement' in such a way as to include these complex sounds. But it is not clear whether the definition may then include other sequences (say [ps] and [ks]) and whether there is really any property of the articulation itself that can settle this question independently of particular phonological systems.

Trubetzkoy's further rules 3–7, which he himself describes as subordinate or secondary to 1 and 2, draw on various other phonetic and systemic criteria, such as the duration of a complex segment in relation to that of simple segments in the given language (rule 3) or the occurrence of elements of a complex segment elsewhere in the language (rule 6; cf. [mb] as a single segment in Fijian, where [b] does not occur on its own, section 3.12 above).

Pike (1947, ch. 12) provides a highly detailed listing of the kinds of phonetic phenomena which are open to interpretation. Under segments which may be either consonant or vowel, he includes not only semivowels but also voiceless vowels (which he suggests need not be syllabic) and a weak velar fricative (which is in effect the semivowel [ɰ]). His list of sequences of two segments which may be interpreted as one includes:

stops with various kinds of release phenomena (aspirated, labialized, affricated, nasally released, etc.);
homorganic nasals plus stops, i.e. prenasalized stops;
vowel glides, i.e. diphthongs;
double stops, such as [k͡p] and [g͡b]; and
sequences of voiced and voiceless equivalents, such as word-final [zs] which may simply be a [z] without full voicing throughout.

He has a separate category of complex segments which may be interpreted as a sequence of two, including long vowels and consonants, vowels with various kinds of secondary articulation (e.g. nasalized [ã] interpreted as [an]) and syllabic consonants (e.g. [m̩] interpreted as [əm]).

Pike's treatment is impressive for its comprehensive attention to phonetic detail and analytical procedure in a field situation. Pike's appeal is to the weight of predominant structural patterning, which he calls 'structural pressure'. More precisely, 'characteristic sequences of sounds exert structural pressure on the

phonemic interpretation of suspicious segments or suspicious sequences of segments' (1947, p. 128). Pike operates with a firm distinction between the phonetic level and the phonological. Thus the terms 'contoid' and 'vocoid' are phonetic terms defined independently of any particular language, while 'consonant' and 'vowel' are (phonological) terms applying within specific languages. For example, an affricate is a sequence of two contoids but a single consonant; a semivowel is a vocoid functioning as a consonant; and so on. Pike's routine, then, is one of assessing phonetic ambiguities in the light of structural patterns. In a language in which all words begin with a single contoid, except that some words begin with [ts] or [dz], the predominant pattern would exert pressure on the linguist to interpret [ts] and [dz] as (single) consonants. This routine has undoubtedly proved useful to field linguists struggling to transcribe and analyse previously unwritten languages. In such situations the linguist often has to decide whether to write [ai] or [aj], [an] or [ã:], and so on.

Some of Pike's exercises and examples may give the unfortunate impression that it is the field linguist's task to impose organization on raw data. Such misgivings are not allayed by Pike's remark to the effect that 'phonetics gathers raw material' while phonological analysis 'cooks' the raw material (1947, p. 57). Nevertheless it is clear from the wider context of Pike's work that he was concerned to analyse language in ways that accorded with native speaker intuitions, and that he was interested in the development of spelling systems that would be efficient from the native speakers' point of view. At the same time Pike was undoubtedly well aware that phonetic raw material is a questionable concept: even highly trained phoneticians cannot transcribe an unfamiliar language in such a way as to provide objective phonetic data for analysis.

It is worth dwelling on this point because linguists such as Pike have been accused of being obsessed with 'taxonomic' analysis and field procedures. In a sense the charge is valid. But it should at least be clear that Pike's 'cooking' of the data is not an exercise in the arbitrary pursuit of regularity or symmetry or notational convenience. Rather, his (and others') work should be judged in the light of the ambition to reduce languages to writing and to correct the natural tendency to interpret all phonological systems against the background of one's own language.

The most serious objection to the formulation of an interpretative routine is that it may give the impression that each doubtful case can be submitted to a decision-making procedure. In fact different criteria may point in different directions, and alternative solutions may simply reflect differing but equally valid perspectives. Frequently, the routine of interpretation is successful only because one criterion, say phonotactic regularity, is pursued to the exclusion of others.

Consider, for example, the [ju] in English *unicorn*, *unity*, *due*, *assume*, *module*. There are at least two reasons for arguing that this complex vowel is a sequence of [j] followed by [u], namely:

1 The [j] can affect a preceding consonant in exactly the same way as any other occurrence of the consonant [j]. Thus the initial consonant of *due* may be [ʤ] while *ss* in *assume* may be [ʃ]. This is parallel to the effect of [j] on a preceding [d] or [s] in sequences such as *could you* and *this year*.

2 Where [ju] begins an English word, we treat it as beginning in a consonant. We say, for example, *a unicorn*, not *an unicorn*.

There are also at least two reasons for arguing that [ju] is a single diphthongal vowel, namely:

3 If [j] is a consonant, then we must allow that English has sequences of consonants such as [dj] and [sj] (as in *due* and *assume*). But these sequences occur only before the vowel [u]. There are no words in English containing sequences such as [dja] or [sji]. This irregularity would not exist if [ju] were interpreted as a diphthong, with [j] understood as part of the complex vowel.
4 The vowel [ju] alternates with the undeniably simple vowel [ʌ]: compare *assume* and *assumption* or *induce* and *induction*. This shift of vowel quality is parallel to other shifts in English: note for example the long vowels in *convene* and *concede* corresponding to short vowels in *convention* and *concession*.

Readers may care to review these four reasons and to consider whether – and why – any of them should take priority over the others.

Exercises

1 How many syllables are there in each of the following words? How would you break each word into syllables? (For example, *banana* has three syllables, *ba-na-na*.)

about, betrayal, camera, categorical, elementary, extra, fascinating, integrity, interesting, sandy, tricky

2 Why is the syllabic structure of some words easier to identify than in others? From the following set of words, list all those which rhyme with each other and all those which alliterate with each other.

cell, coal, fed, fell, head, heed, keen, keyed, king, said, scene, seed, sing, soul, whole

3 Check that you understand the meaning of each of the following terms.

a. double stop
b. homorganic
c. labialization
d. phonotactics
e. semivowel
f. simultaneous palatalization
g. sonority

h. velarization
i. vocoid
j. vowel retroflexion

4 What is an affricate? Why are the sounds at the end of the English words *catch* and *badge* usually understood to be affricates, but not those at the end of *cats* and *adze*?
5 What is a diphthong? Give at least two examples from English.
6 What makes a consonant syllabic? Give examples of syllabic consonants from as many languages as possible.
7 Explain the difference between nasalization of a vowel and prenasalization of a plosive.
8 Why is it simplistic to suppose that speech consists of individual sounds put together in sequence?

4 The Phonemic Organization of Speech

Introduction

This chapter explores a long-standing and fundamental insight into spoken language – that it can be understood as the realization of a system of phonemes. The chapter begins by placing the phoneme in the context of the inherent variability of speech (4.1). It then explains and illustrates what is meant by 'phoneme' (4.2) and by the related concept of 'allophone' (4.3).

This basic introduction is followed by a series of topics which are a necessary part of conventional phonemic description but which also need to be addressed as theoretical issues:

- the notion of phonemic norms (4.4)
- pattern and symmetry in phonemic systems (4.5)
- the question of phonological reality (4.6)
- the relevance of units and boundaries in speech (4.7)
- phonemic invariance and overlap (4.8)
- biuniqueness in phonemic analysis and the neutralization of phonemic distinctions (4.9)
- morphophonemic alternation (4.10)
- free variation (4.11).

The chapter ends with a review of the kinds of phonemic systems that are found across the languages of the world (4.12).

4.1 Phonetic variability

In chapters 2 and 3 we have seen how various articulatory gestures and processes can be used to generate speech sounds and how particular languages organize the flow of speech within structured patterns. Putting it very simply, we can say that a language selects from the human articulatory potential, and that it systematizes that selection. In consequence individual languages (and