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

dialects) are normative, in the sense that speakers operate within the limits imposed by such selection and systematization. This phonological normativity is not of course a matter of legal obligation or moral duty, nor in most cases does it emerge from formal training or instruction in pronunciation; rather it unfolds in the process of our growing up in a particular speech community, and acquiring and maintaining the speech habits of that community. We show our response to such normativity in dozens of ways – often quite informally or even subconsciously – whenever we identify a particular pronunciation as strange or foreign, when we recognize and warm to a familiar regional dialect, or when we dismiss a foreign word or name as 'unpronounceable'.

This is not to imply that we are all loyally attached to a single local dialect or language. Speakers of a language such as English, spread across a large and diverse population around the world, may be familiar with many different norms and may themselves exploit different norms according to circumstance, shifting, say, between a local or informal style of pronunciation and one that would be considered more standardized or formal. Nevertheless, while such versatility may complicate the status and application of phonological norms, it does not deny the existence and strength of the norms themselves.

If we do not acknowledge this normative character, we have little justification for talking about and investigating 'normal' pronunciation. If we do acknowledge it, we have a basis at least for describing pronunciation against a background of what counts as normal. For pronunciation is in fact highly variable, even within the limits of what may be agreed as normal. This is hardly surprising, given the nature of the articulatory mechanism, the precision and coordination needed to control it, and the fineness of auditory discrimination.

A problem for phoneticians, but not for the average user of language, is that there are considerable physical differences among speakers. Variations in the size and shape of the vocal tract and articulators are sufficient to yield substantial and persistent differences between one speaker and another. (As we shall see later, in chapter 7, it is a challenge to explain how it is that we can discount such differences – how we manage to hear 'the same words' being uttered by two quite different speakers, and yet at the same time respond to the differences by identifying the two voices as particular individuals.) An obvious and striking example is the difference between a child and an adult in, among other things, the overall length of the vocal tract. The difference in length – far greater in adults than in children – has major effects on the sound quality of speech, yet we are able to allow for this in our hearing of children's speech.

Another relatively permanent cause of differences is that individuals learn or become accustomed to habitual settings in the underlying postures of articulators – in much the same way that individuals have habitual body postures of which they are barely conscious but which affect the way they characteristically sit and stand and walk. There are also wide variations in habitual rate of articulation, and differences in the laryngeal settings used for 'normal' voiced phonation. A speaker may, for example, always use somewhat breathy phonation, or always articulate with the lips slightly protruded, or always use a relatively slow rate of articulation. Such differences usually do not affect the articulation of individual speech sounds in a particular or selective way,

but are global properties that contribute to a total impression of voice quality. (See Laver 1980 for a comprehensive discussion of the phonetics of voice quality.)

Apart from these global differences among individuals, many speakers have a characteristic way of articulating certain sounds. For example, a particular speaker of English may, regularly and systematically, produce alveolar plosives with unusual fronting, almost as dentals. This is likely to be a noticeable feature of the individual's speech, the kind of thing a mimic might fasten on to. Among English speakers there are sizeable minorities who pronounce [s] and [z] sufficiently unusually to be noticed (and sometimes to be described as 'lisping') or who use an r-sound with a high degree of lip protrusion (which may lead to the accusation that they 'say *w* instead of *r*').

Variability such as we have mentioned so far is often described as pervasive. But speakers may also vary their articulatory behaviour, consciously or unconsciously, in a way which is often unpredictable and certainly not pervasive. This sort of idiosyncratic variation may often go unnoticed or be dismissed as trivial oddity, and it is generally tightly constrained by the demands of the phonological system. Thus a speaker of English who happens in one particular utterance to devoice the initial consonant of the word *zip* is likely to be heard as having said the wrong word, namely *sip*. While context may make it perfectly clear that *zip* was intended, a systemic error of this kind is more likely to attract attention than, say, devoicing of the [z] in *adze* or *adds*, where both words are normally pronounced identically and the voicing is not distinctive. In general, the phonological system of any language will make some variations far more tolerable than others.

Certain aspects of speech may vary according to the speaker's social environment and emotional state. Speakers will generally exercise considerably more articulatory care when making a speech on a formal occasion than when chatting casually with friends. Lindblom (1990) formalizes this idea in what is now commonly referred to as the hyper–hypo articulation continuum, i.e. the continuum along which speakers adapt their articulation strategies to the communicative needs of the situation. The articulatory consequences of such deliberate attention to speech cannot always be easily distinguished from the involuntary effects of the speaker's emotional state. Anxiety or fear or anger can noticeably affect articulation rate, phonation mode or articulatory forcefulness, and we are all accustomed to reading emotions from an overall impression of these properties of speech. Similarly, articulation may change quite radically as a speaker makes special efforts to be heard intelligibly in adverse circumstances, such as against a background of noise. Effects such as these are often described as 'affective' or 'paralinguistic', implying that they are a matter of general background, peripheral to the main communicative function of language, but it is in fact not at all easy to quantify and predict these factors in such a way as to separate them off from 'truly linguistic' functions. Consider, for example, the difficulty of distinguishing between anger as a communicative strategy – with features of articulation deliberately adopted and under control for persuasive or threatening purposes – and anger as an uncontrolled and involuntary emotion. In any case, so-called paralinguistic features do contribute

significantly to variability in articulation, both within the speech of an individual and from one speaker to another.

Traditionally more central to linguistic description is CONTEXT-SENSITIVE VARIATION. Speech does not consist simply of a string of target articulations linked by simple movement between them (section 3.1 above). Instead, the articulation of individual segments is almost always influenced by the articulation of neighbouring segments, often to the point of considerable overlapping of articulatory activities. As a consequence, the notional or 'ideal' way of articulating a particular sound is subject to modification in running speech. This phonetic variability is due not just to differences among individual speakers, but also to the phonetic context. The general effect is known as CONTEXT-SENSITIVITY.

Context-sensitive variation has complex and interacting causes which are not yet completely understood. Two basic types can be distinguished: (1) the effects of the biomechanical performance properties of the vocal tract, and (2) the effects of the nature and organization of the neuromuscular control mechanisms which actuate articulator movements. Both types may reflect genuine limitations on what the vocal tract can achieve – there are, after all, limits to the speed with which the tongue can move from one position to another, or to the rate at which the vocal folds can vibrate. But the other side of the coin is that both types may reflect the level of articulatory performance that is sufficient to produce adequate phonetic distinctiveness in the language in question. In many instances what is required for the language makes it irrelevant to ask what the limits of articulatory potential are – linguistic organization is such that articulation does not, so to speak, stretch the machinery to its limits. It is context-sensitivity that accounts for much of the complexity and indirectness in the relationship between the acoustic output of articulatory activity and the linguistic structure which it represents. As a result, the way in which linguistic structure is encoded in the acoustic speech signal is rather opaque. Despite that, listeners can decode it with apparently unconscious ease.

Nevertheless, the pressure of context may have quite noticeable effects. We do not normally think of English as having nasalized vowels – in the way that French and Portuguese have a distinction between oral and nasalized vowels. But vowels preceding nasal consonants in English, as in *sand* or *can't* or *bend*, may well be nasalized because of the following consonant. Even more radically, many English segments may be articulated in certain contexts as sounds from which they are normally distinguished. Thus [s] is distinct from [ʃ] in English, but [s] may nevertheless be articulated as [ʃ] in an appropriate context, as for example when the [s] immediately precedes [j], as in *this year* or *tissue*. While effects of this kind may still go unnoticed if they are common enough in the community, they may also attract attention, especially if there is a division between speakers who tolerate the context-sensitivity and others who try to suppress it.

The causes and mechanisms of context-sensitivity have been the subject of a fair amount of research, contributing to our understanding of articulatory dynamics and raising new questions about the high-level neural representation and organization of muscular commands and the transformation of these commands into articulatory movements. The vocal tract, including the articulators

within it, forms a biomechanical system which is subject to the laws governing all mechanical systems, from can-openers to space shuttles. Specifically, the mass and size of articulators constrain their movement in relation to the muscle systems that actuate them. Articulators have mass and are subject to inertia: they resist being set in motion. There is therefore some inherent delay between a neuromuscular command and the intended articulatory gesture. The greater the mass, the greater the inertia and hence the greater the delay.

A common example of this effect is the tendency for peripheral vowels in short syllables to become centralized, particularly when the speaker is talking rapidly. In simple terms, the tongue may not have time to reach the target position before the next sound has to be articulated. While the tongue is moving towards the peripheral target position determined by the neuromuscular commands, conflicting commands for the following segment are already arriving, initiating movement towards a different position. The result of this conflict is a general tendency for the tongue to assume a more central or neutral position, effectively smoothing or summing the mechanical consequences of the individual movement commands. In effect, the average or long-term 'context' of tongue position is central, and biomechanical inertia heightens the tendency to centralization as the speaker attempts a faster rate of movement. But this tendency is not just a matter of yielding to the constraints of biomechanical performance, for the speaker may also impose limits on the muscular activity used to overcome mechanical inertia. In other words, a speaker may, to varying degrees, either make efforts to operate the articulatory system to the upper limits of its performance or lower the performance to accommodate to the system. Whatever its cause, the effect is known as target UNDERSHOOT: the principal articulator fails to reach the target position defined in the canonical description of the segment. The centralization of peripheral vowels by undershoot is commonly known as VOWEL REDUCTION (Lindblom 1963, Stevens and House 1963, Stevens et al. 1966, Delattre 1969, Moon and Lindblom 1994).

The effect of delay on articulator movement can be seen in English words such as *more* and *now*, where the (beginning of the) vowel is nasalized, partly because of delay in raising the velum at the end of the nasal consonant. The nasality of the initial consonant thus overlaps on to the following nominally oral vowel. A similar effect tends to nasalize the voiced fricative following [n] in words such as *burns* and *bronze*. Comparable effects of delay can be observed in words such as *paws* and *jaws*, where the lip rounding of the vowel is likely to persist into the alveolar fricative at the end of the words. By comparison, the same fricative has spread or neutral lip position in words such as *bees* or *haze*.

The organization of neuromuscular commands may also produce the very opposite effect. To compensate for inherent delay, neuromuscular commands may be initiated well before the segment for which they are required; articulatory properties of that segment may then appear on an earlier segment. This, then, is an anticipatory form of overlap. A simple example is provided by the nasalization of the vowel in words such as *sand* and *can't*, where the velum may be lowered during the vowel in anticipation of the following nasal consonant. Anticipation likewise affects the point of articulation of velar plosives in English, in words like *key*, *car* and *core*: the stop closure is most forward

in *key* and most retracted in *core*, because the tongue body anticipates the position required for the following vowel. A third example is the lip rounding on alveolar fricatives in words such as *saw* and *sue*, which anticipates the demands of the following rounded vowel.

Amerman and Daniloff (1977) have shown that when a speaker articulates a CCV sequence, the tongue body may begin to move towards the vowel even during the first consonant. Similarly, in VCC sequences, anticipatory movements towards the second consonant can start during the vowel. According to Benguerel and Cowan (1974), lip protrusion may be evident several consonants in advance of the rounded vowel for which it is required, while Amerman et al. (1970) note that speakers may likewise anticipate a relatively open vowel by beginning to lower the jaw during preceding consonants. A number of researchers have also found that in the first vowel of a VCV sequence, there may be tongue movement anticipating the second vowel across an intervening consonant (e.g. Öhman 1966, Butcher and Weiher 1976, Fowler and Brancazio 2000, Recasens 2002).

These context-sensitive effects underline the danger of assuming that individual segments (and their articulatory properties) have any real autonomy within connected speech. Features of articulation interact and overlap, in both anticipatory and perseverative fashion, sometimes extending over several segments.

Context-dependent overlap of the kind we have been describing is often known as COARTICULATION. The reader should note, however, that this term is not used consistently. Some writers use it in the narrow and rather literal sense of simultaneous movement of two different articulators. Under this definition, the lip rounding of a consonant, anticipating the rounding of the following vowel (as in *saw* or *sue*), is coarticulation, but the adjustment of the tongue position for a velar consonant, anticipating the tongue posture of the following vowel (as in *key* and *core*), is not. The second kind of phenomenon may be described as 'adaptation' or 'accommodation' – the articulator (in this example the tongue) is, so to speak, reaching a compromise with the demands of an adjacent articulation. Our own usage is to describe both types of context-dependent overlap as coarticulation, without reference to the number of articulators involved.

Perseverative coarticulation effects are known as LEFT-TO-RIGHT COARTICULATION (in short, L > R). Thus in the string . . . AB . . . , sound A influences sound B (or beyond). L > R coarticulation is thought to be largely due to lag in articulatory movement, induced by inertia. The relevant ingredients are the biomechanical properties of the articulators (their size and mass, and the nature of the muscles involved); the speaker's rate of articulation; and the extent to which the speaker is exercising voluntary neuromuscular effort in the control and movement of the articulators.

Anticipatory coarticulation effects are known as RIGHT-TO-LEFT (L < R) COARTICULATION. In the string . . . CD . . . , sound D influences sound C (or earlier sounds). L < R coarticulation is thought to be due to deliberate high-level organization of the neuromuscular commands for the relevant sounds. This high-level planning is complicated by the differences in innervation latencies among the various articulatory muscle systems.

Again, if we think of speech as a series of autonomous segments, we are in danger of dismissing coarticulatory overlap as a sort of needless complication, interfering with the ideal properties of speech. But, on the contrary, coarticulation is an essential characteristic of speech. Speech production depends on very rapid, highly coordinated articulatory movements, and it is doubtful whether we could achieve anything like the articulation rates of normal running speech if we did not make extensive use of overlap. Daniloff (1973) claims that the tongue tip – the fastest of the articulators controlled by muscles – can perform only about eight closures per second. We are nevertheless able to produce from 12 to 18 segments per second in running speech. Thus coarticulatory overlap enables us to work very effectively within the constraints on our performance. As a consequence, the quasicontinuous fluidity of speech can be thought of as efficient encoding, rather than as degradation of the signal.

It is not always easy to determine how far speakers are simply constrained by the limits of the biomechanical system and how far they are actually setting a level of articulatory performance that is just sufficient to meet the demands of their language – just sufficient, that is, to be adequately intelligible in the immediate circumstances. Lindblom (1983) argues strongly that distinctiveness and communicative effectiveness are primary motives in speech production. It is certainly true that general tendencies such as vowel reduction and anticipatory nasalization are not uniform in their effect on different languages. For example, although it is generally true that a faster rate of articulation is likely to increase the amount of vowel reduction, a comparison of speakers of English from different parts of the world would show different responses to this tendency. At a given rate of articulation, speakers of RP probably reveal appreciably more examples of reduction than, say, northern English or Australian speakers, well before any biomechanically imposed limit is reached. Consider, for instance, the variability in words such as *hostel*, *synod* and *bursar*, where the second vowel may or may not be reduced: the choice of the reduced form is likely to be influenced more by the speaker's sense of a correct or natural pronunciation within the relevant community than by rate of articulation. (More general observations about the reduced vowel in English can be found in Gimson 1980, pp. 126–7 and 224–5). A similar point could be made about nasality, since the occurrence and extent of both anticipatory and perseverative nasality varies considerably among speakers and languages. Such examples suggest that language-specific phonological norms and patterns play a major role in determining the nature of speech. The term PHONOLOGICAL CONDITIONING is widely used to explain variability which seems to be a matter of language-specific 'rules of pronunciation' (section 4.3 below).

The term ASSIMILATION has a longer tradition than coarticulation, and is sometimes used in a rather general way, more or less synonymously with coarticulation. Quite often the term refers only to those cases of context-sensitive articulatory overlap which are reflected in phonetic transcription. In this usage, the term becomes rather too dependent on ill-defined conventions about the nature of transcription. Thus assimilation may include instances of overlap which happen to generate a change from one common sound to another (as when the alveolar [n] of *un-* becomes velar [ŋ] before a velar plosive in *unkind*

or *ungainly*), but exclude instances that give rise to a less common sound for which there is no well-known phonetic symbol (as when the initial consonants of *saw* or *sue* are lip-rounded in anticipation of the following rounded vowels). In other words, what counts as assimilation tends to depend on the availability of symbols to indicate it and on conventional judgements about its auditory or linguistic salience. Many effects, such as changes in the tongue body posture of alveolar stops in the context of different vowels, are not even allowed for in conventional phonetic transcription, and so are likely to be ignored in accounts of assimilation.

Assimilation is often mentioned in connection with historical changes, and many of the sound changes that languages have been observed to undergo can appropriately be described as assimilatory. Thus English words such as *mission*, *passion*, *special*, *crucial*, *nation* and *lotion* were once pronounced with a medial [sj] or [si] but in modern English have [ʃ]: by a process of assimilation, the [s] has been retracted in anticipation of the following [j] or [i] (which has then disappeared, or been 'swallowed up' in the assimilatory process). We must, however, distinguish between historical processes and processes that are still current or operative in the modern language. We know, for example, that words such as *ship* and *shall* are derived from older forms (in Old English or even earlier) beginning with [sk]. Here a sound change has had its effect on the language, and we have no access to the earlier pronunciation other than by historical investigation and comparison with other related languages. (Part of the evidence for the change, for example, is that Old English records reveal the spellings *scip* and *sceal*, while other old Germanic languages, notably Old Norse and Gothic, show *skip* and *skal*.) Note that this change is in a real sense over and done with. There is no tendency in modern English speech to repeat the process in words such as *skill* or *sky*, for instance. On the other hand, there are processes which can be observed within the current state of the language. The assimilation of alveolar [n] to velar [ŋ] before velars, for instance, is demonstrable within modern English. There are forms such as the prefix *un-* which clearly have [n] in nonvelar contexts (*untidy*, *unsettled*, etc.) but which may have [ŋ] before a velar (*unkind*, *ungainly*, etc.); and the process can be seen to apply to many words that normally have alveolar [n], as when unstressed *can* precedes a word beginning with a velar (*they can* [ŋ] *keep it*, *you can* [ŋ] *go now*) or when words like *pan* and *sun* are compounded in *pan*[ŋ]*cake* or *sun*[ŋ]*glasses*. In describing the system and structure of pronunciation in the current language, we need take no account of historical changes that are over and done with; indeed, it would be inappropriate to do so, for from the point of view of a speaker of the modern language, these changes have disappeared over the horizon. But assimilatory processes that can be observed at work within the modern language certainly are part of the modern speaker's organization of pronunciation and are relevant to our description of the language.

It is also important to note here that English spelling, taken without other evidence, is no sure guide to either historical or current processes of assimilation. Thus it happens to be true that the spelling *ssi* in *mission* or *passion* suggests an earlier pronunciation with [si]; but the spelling *sh* in *ship* and *shall* does not indicate a previous pronunciation as [s] followed by [h] (as in *mess-hall* or *doss-house*). Moreover, English spelling abounds in oddities that make it quite unreliable in this regard: for example, the *l* in *should* and *would* is indeed a pointer to an earlier pronunciation with [l], but the *l* in *could* is there by analogy with the other two forms, and the word has never been pronounced with [l]. If we want to demonstrate relationships among sounds in the modern language, we must appeal not to spelling but to pronunciations that can be recorded, checked and compared. Thus the spelling of *mission* is in itself no reason to connect the [ʃ] of its pronunciation with an [s]. But we can show a relationship between the [ʃ] and [s] by appealing to the forms *submission* and *submissive* or *permission* and *permissive*. (In fact appeals of just this kind are central to the generative approach to phonology, which we shall outline in chapter 5.)

Traditional use of the term assimilation focuses on the more obvious or more easily symbolized consequences of coarticulatory effects, and for this very reason the term is widely known, especially with reference to consonants. In a non-technical way, three types of assimilation can be identified. ASSIMILATION OF PLACE is exemplified by English *ratbag* or *oatmeal* pronounced with [p] instead of [t] in rapid or informal speech, by assimilation of the alveolar stop to a following bilabial. ASSIMILATION OF MANNER refers to instances such as *Indian* pronounced as *Injun*, where the stop [d] and approximant [j] merge to form an affricate. (While *Injun* is generally considered substandard in modern English, the same assimilation has applied historically in *soldier*, in which the affricate is now normal.) ASSIMILATION OF VOICING is illustrated by *have to* pronounced with [f] rather than [v], by assimilation of the voiced fricative to a following voiceless consonant.

ELISION refers to the special case of loss or omission of segments or syllables. Sounds may be so weakly articulated that they no longer have auditory significance, or they may be omitted altogether in the stream of running speech, particularly – but not exclusively – in casual or rapid speech. Like other phonetic variations we have looked at, elision is constrained by the phonological system and often applies to segments and weakly stressed syllables whose absence does not seriously impair intelligibility for native speakers of the language. In English, elision is often found in consonant clusters, as in *facts* and *chests* pronounced without [t], or *fifths* and *sixths* pronounced without [θ]. When unstressed, the word *and* often loses the [d], and an entire unstressed syllable is often elided from longer words such as *February* and *library*. In many languages, word-final unstressed vowels may be elided, either in general or when the next word begins with a vowel. In French, instances of such elision are standard and are marked in orthography by an apostrophe, as in *j'ai* 'I have' or *l'air* 'the air', where an unelided *je ai* or *le air* would be simply incorrect.

The question of context-sensitive effects and their causes continues to cause lively debate among speech researchers. We conclude this section with a conservative summary of what is known. In the first place, coarticulation effects seem capable of spreading across several segments, and are often not checked unless they are in direct conflict with other articulatory demands, or unless they run up against the contrastive requirements of the language. Secondly, observable assimilations seem to be caused more often by anticipatory coarticulation effects than by perseverative effects, at least in English if not in most

languages. Thirdly, even if we restrict our attention to vocal tract performance alone, coarticulation effects are not yet fully understood: it is not clear to what extent we can explain them by assuming that high-level commands associated with specific segments are confounded by biomechanical 'sloppiness' and the unequal latencies of the neuromuscular innervation system; or to what extent high-level commands are quite deliberately planned to optimize transitions between targets and to yield the best possible vocal tract performance in running speech. In general, it does seem that the limitations of vocal tract performance are not predominant in influencing context-related variability. Fourthly, there is often no simple way of distinguishing between those assimilation effects which are due to the inherent properties or limitations of speech production and those which are not, unless the latter are very obviously language-specific. Assimilation often appears to be motivated by ease of articulation, but what seems easy and natural in one language often turns out to be less so in another. Thus ease of articulation needs to be assessed within the constraints of differing languages, each with its own system and structure.

4.2 The phoneme

The constant background to our discussion of variability in the previous section has been the observation that in any language some differences in pronunciation are crucially distinctive. It is these distinctions or contrasts that are recognized by speakers of the language as 'making different words' and acknowledged by linguists as systemically functional. In English, for example, we must differentiate words such as *led*, *red* and *wed* from each other if we are to achieve acceptable pronunciation; and similarly *allay*, *array* and *away*, and *click*, *crick* and *quick*. Abstracting the individual sounds from the normal flow of speech, we can say that in English the three consonants *l*, *r* and *w* are CONTRASTIVE or DISTINCTIVE.

The phonological system of English is such that each of these sounds may vary considerably in its articulation. The *r* in *tree* and *train* may be a voiceless fricative, the *r* in *dream* and *drain* a voiced fricative, and the *r* in *three* and *throw* a tap or flap, all three of these variants being phonetically quite different from the *r* in *red* or *array*. Not all speakers of English pronounce *r* in the same ways, of course, but the general point is that what counts as a single sound within a system may be articulated in various ways provided that contrasts are maintained (i.e. provided that *train* is still distinct from other words such as *twain* or *chain*, and that *drain* is still distinct from *Jane*, and so on).

To avoid any misunderstanding about the 'English phonological system', we should stress that it is actually not one system but many, for dialects as well as languages can differ in their system of phonological contrasts. In English, though certainly not in all languages, it is the vowel contrasts that differ most; readers may care to check their own pronunciation of the following words, arranged in five columns:

(1)	(2)	(3)	(4)	(5)
spa	spar	saw	spore	spoor
Pa	par	paw	pore	poor
Ma	mar	maw	more	moor

For some speakers of English (including the authors) only two contrastive vowels are represented here, a rather central long [aː] in columns (1) and (2), and a rather more back and rounded [ɔː] in columns (3), (4) and (5). This is a version of English in which final *r* is not pronounced, and in which the words adjacent to each other in columns (1) and (2) are therefore identical. Those who do have a final *r* (many American and Scottish speakers, for example) may distinguish the vowel of column (4) from that of column (5); on the other hand, at least some of these speakers may use the same somewhat rounded vowel for columns (1) and (3). Thus the number of contrasts, as well as the nature and variability of individual sounds, may certainly differ from dialect to dialect within a language.

The extent to which variant pronunciation counts as 'saying the same sound in a slightly different way' will obviously depend on the linguistic system. A number of the world's languages (including Classical Arabic and some Australian Aboriginal languages) have only three contrastive vowels, which can be represented as *i*, *a* and *u*. In such languages, the quality of the *a* vowel may vary considerably, say from a back rounded [ɒ] in the neighbourhood of consonants such as [w] to a front [æ] in the neighbourhood of [j] or other palatal consonants. Such variation cannot be systematically tolerated in a language in which [ɒ] and [æ] are distinct phonemes.

Contrastive systems range in complexity from languages with fewer than 20 distinctive consonants and vowels to languages with 60 or more. English, depending on the particular dialect, has up to 24 consonants and up to about 20 vowels. English has a rather high number of vowel contrasts, especially in comparison with a typical Australian Aboriginal language. On the other hand, most Aboriginal languages have a contrast between at least two and sometimes three kinds of r-sound. For example in Warlpiri, from central Australia, we have:

marru house	*rr* represents trilled [r]
tjarra flame	
maru black	*r* represents approximant [ɹ]
tjara fat	
mardu wooden bowl	*rd* represents retroflex flap [ɽ]
tjarda sleep	

It is difficult to formulate comparisons of this kind without adopting the perspective of one particular language system: we are inclined to say that Warlpiri has 'three r-sounds' but from the Warlpiri point of view the three sounds are not three versions of one sound, but three distinct consonants, as crucially different from each other as *l*, *r* and *w* are in English.

A common way of conceptualizing such phenomena in modern linguistics is through the notion of the PHONEME. Although the notion remains controversial,

it rests ultimately on the recognition of functional differences. English speakers take *led* and *red* to be different words, as Warlpiri speakers take *marru* and *maru* to be different words. A phoneme can thus be described as a contrastive or distinctive sound within a language. [r] and [ɹ] and [ɽ] are separate phonemes in Warlpiri but not in English; [ɒ] and [æ] are separate phonemes in (most varieties of) English but not in Warlpiri.

Sounds which count as alternative ways of saying a phoneme may be termed VARIANTS or ALLOPHONES. A common convention is to use slant lines to indicate phonemes and to retain square brackets for the phonetic notation of allophones, e.g.

English /r/ may be realized as [r], [ɹ], etc.
Warlpiri /a/ may be realized as [ɒ], [æ], etc.

For any of the world's languages, then, it is possible to draw up an inventory of phonemes, each of which will have one or more variants or allophones. Although this will by no means exhaust what can be said about the phonological system of a language, it will in effect be a list of the significant or contrastive sounds of the language with a specification of major phonetic variants for each phoneme.

In most cases, allophones will fairly evidently be governed by the processes or patterns of the language concerned (many of them due to coarticulatory effects of the kind discussed in section 4.1 above). This implies that each allophone occurs in a particular phonetic environment or specifiable context. The phoneme /n/ in English, for example, may have three allophones as follows:

Phoneme	*Allophones*
/n/	[n̪] before a dental fricative
	[n:] before a voiced obstruent in the same syllable
	[n] elsewhere

Thus /n/ is dental by assimilation in e.g. *tenth* or *month*, and is lengthened before [d], [z] or [ʤ] in e.g. *tend*, *tens* or *lunge*. Where neither of those two conditions applies, the phoneme has its 'normal' English value of [n], as in *net*, *ten* or *tent*.

An inventory of phonemes can be viewed in two directions. Seen from the point of view of the language system, it represents those sounds which are significant in the language: the phonemes are those sounds which serve to differentiate words. From this perspective, what matters about a phoneme is not so much the precise ways in which it may be pronounced but rather the fact that it is different from the other phonemes of the language. Hence the importance attached to pairs of words differing in only one phoneme, such as English *red* versus *led*, *red* versus *wed*, *real* versus *zeal*. These pairs, known as MINIMAL PAIRS, provide solid evidence of phonemic contrasts, of the differences that matter in a language, and they are of interest not just to the phonological analyst but also in such fields as language teaching and hearing testing.

We can, however, also view phonemes from the point of view of their actual pronunciation. In this case we are, so to speak, looking upwards from the level of a narrow phonetic transcription. Seen from this angle, a phoneme is a set of related sounds or phones. Allophones are similar sounds occurring in complementary environments: English [n̪] is found only before a dental fricative, never in any other environment; lengthened [n:] occurs only before voiced obstruents and never elsewhere; and so on. Hence, where a phoneme has more than one variant, it may be said to consist of a set of allophones standing in COMPLEMENTARY DISTRIBUTION or in MUTUALLY EXCLUSIVE ENVIRONMENTS.

These two perspectives on the phoneme have sometimes been set against each other as, say, a 'functional' view of the phoneme as opposed to a 'phonetic' view of the phoneme. We take the two notions of the phoneme to reflect two different aspects of the same phonological reality. To the native speaker, this reality means on the one hand, 'functionally', that certain differences in pronunciation are genuine or real; and on the other hand, 'phonetically', that a good deal of phonetic variability may be tolerated within a phoneme.

4.3 Allophones

In general, allophones can be described as CONDITIONED variants of a phoneme, generated by PHONOLOGICAL CONDITIONING. Phonological conditioning is usually understood to be a matter of language-specific 'rules of pronunciation', although we have already noted that it is often difficult to draw a clear boundary between the effects of the biomechanical system and the effects of the linguistic system (section 4.1 above). There are observable universal tendencies in pronunciation, but languages differ enormously in the extent to which they constrain or suppress these tendencies. Moreover, some instances of phonological conditioning have little or no apparent biomechanical justification. In such instances, the habitual pronunciation of a language may be strikingly odd to speakers of other languages (and far from easy for others to imitate). Certainly, phonemic analysis customarily describes as allophones only those major variants that can be categorized and represented in a segmental transcription, and these tend to represent variation which is not universal, even if found in a substantial number of languages; variability that is revealed only by instrumental analysis is ignored. Some examples follow.

/a/ [ã] before a nasal consonant
[a] elsewhere.

Redundant or nonsignificant nasalization of vowels is observed in many languages, including at least some varieties of English (nasalized vowels in *can't*, *sand*, but not in *cat*, *cart*, *sad*).

/k/ [g] between two voiced sounds
[k] elsewhere.

ditioning of this kind occurs in many Australian Aboriginal languages and
r languages in which there is no phonemic contrast between voiced and
eless sounds: the plosive is voiced in a fully voiced context but not other-
(e.g. not in word-initial position).

[ŋ] before a velar consonant
[n] elsewhere.

applies for instance to Italian and Spanish, in which there is no phoneme
Wherever /n/ immediately precedes /k/ or /g/ it is assimilated to the velar
:ion, e.g. in words such as *banca* and *mango*. A comparable assimilation
und in English, with a velar nasal preceding the velar stop in words such
nk, *bank*, *anger*; but /n/ and /ŋ/ are in contrast in English in minimal pairs
as *sin/sing*, *run/rung*, *sinner/singer*, etc. There are, however, varieties of
ish in which this contrast does not exist, namely those in which words
as *sing*, *rung*, *singer* are pronounced with [g] following the velar nasal
sing is [sɪŋg]). These varieties of English (chiefly found in the Midlands
ıorth of England) are like Italian and Spanish in that [ŋ] occurs only immedi-
before a velar consonant and can therefore be analysed as a conditioned
nt of /n/.

[ð] between two vowels
[d] elsewhere.

is instance the plosive is 'weakened' or 'lenited' to a fricative when between
ls. A process of this kind is observable in Spanish and Portuguese, where
nedial *b*, *d*, *g* in words such as *Cuba*, *Toledo* and *Diego* are generally
ılated as voiced fricatives rather than as plosives.
›cesses of conditioning are not always obvious from a segmental transcrip-
imagine that the following are words of a language, phonetically transcribed:

ı]	[komo]	[komu]	[mini]	[mito]	[mɵki]
o]	[nipu]	[nyti]	[piti]	[puko]	[pymi]
	[tonu]	[tøni]	[tɵpi]	[tunu]	[tyki]

that the vowels are high front unrounded [i], high front rounded [y],
ront rounded [ø], high back rounded [u] and mid back rounded [o]. Now
ıd [u] are in complementary distribution, as are [ø] and [o]. [y] occurs
in the first syllable where the following syllable contains [i], whereas [u]
s only in other positions, namely in the second syllable or in the first
le if the following syllable does not contain [i]. A parallel distribution
e noted for [ø] and [o]. The phonetic explanation is that the high front
on of [i] is anticipated in the preceding vowels [u] and [o], which are
·d to [y] and [ø]. The vowel phonemes of this artificial language are thus

[i]
[y] before a syllable containing [i]
[u] elsewhere

/o/ [ɵ] before a syllable containing [i]
[o] elsewhere

Phenomena of this general type are known variously as VOWEL HARMONY, UMLAUT or MUTATION, and are a vivid demonstration of coarticulation effects. The German term *Umlaut* and its English equivalent 'mutation', taken in their narrow sense, refer specifically to certain processes operative at earlier stages of Germanic languages. The difference in vowels in English *foot/feet*, *mouse/mice* or German *hoch/Höhe* ('high/height'), *Kuh/Kühe* ('cow/cows') is in fact due to a process of precisely the kind exemplified in our artificial data. Both English and German have undergone subsequent sound changes which have obscured the original conditioning, but, inasmuch as earlier pronunciation can be reconstructed, *foot* and *feet* are derived from something like [fo:t] and [fø:ti], *hoch* and *Höhe* from [ho:χ] and [hø:χi].

In most of the above examples it is relatively easy to point to CONDITIONING FACTORS, features of the context that are responsible for the allophonic variation – the nasal consonant that conditions nasalization of a preceding vowel, the voicing of vowels that conditions voicing of an intervocalic stop, and so on. In these cases, the processes affecting the phonemes seem general or 'natural' tendencies of speech. But it is obviously not true that these tendencies yield identical consequences in all languages. Furthermore, some instances of allophonic variation are relatively difficult to explain in phonetic terms, and it is not at all easy to find plausible conditioning factors. Some examples follow.

In Korean, [l] and [r] are allophones of one phoneme, with [r] standing word-initial and between two vowels, and [l] elsewhere. The notion that [l] and [r] are really 'the same sound' is of course quite contrary to the expectations of speakers of many other languages. The 'similarity' of [l] and [r] is not easy to justify, although it is worth noting that even in a language such as English, in which *l* and *r* are distinct phonemes, the two consonants are prone to confusion; witness the way in which even fluent native speakers may stumble over words containing *l* and *r* in 'awkward' combinations, e.g. *meteorological*, *corollary*, *irrelevantly*, etc. The similarity that allows these two consonants to be identified or confused must be understood systemically: in many languages *l* and *r* are the only two continuant consonants which are neither fricative nor nasal.

In a few Australian Aboriginal languages, a lamino-dental stop and a lamino-palatal stop are allophones of a single phoneme. Although both consonants are articulated laminally (with the blade rather than the tip of the tongue), the auditory effect is quite different, at least to those who are not native speakers. The details are often complicated (see e.g. Glass and Hackett 1970, pp. 109–10, for a description of what happens in one dialect of Pitjantjatjara or Western Desert) but the general pattern is that lamino-palatal [c] occurs before a front vowel [i], whereas lamino-dental [t̪] occurs before other vowels, i.e. [a] and [u]. While it is normal for vowel quality to cause some kind of modification to a preceding consonant, variation from dental to palatal articulation is unusual among the world's languages.

In standard Indonesian, the phoneme /k/ has a glottal stop allophone occurring word-finally, as in *duduk* ('sit'), *tarik* ('pull') pronounced with final [ʔ].

While the adjustment can be explained as substitution of a glottal closure for the velar closure of a [k], it is certainly not a substitution that comes easily and naturally to speakers of most other languages.

Finally, as an instance of an allophonic adjustment which happens in many varieties of English but is far from universal, we note the distinction between clear and dark laterals. (The 'dark l' is velarized by raising of the back of the tongue towards the soft palate; see section 3.6 above.) The clear variant normally occurs before a vowel (*lend*, *alight*, *believe*) and the dark before a consonant or word-finally (*wild*, *halt*, *will*, *hall*). The velarization is extreme in some varieties of English, notably in the speech of many Londoners and South Australians, who may even fail to make the lateral occlusion. As a result, the raising of the back of the tongue virtually creates an [u] vowel (cf. *hall* pronounced as [hɔ:u], *halt* as [hɒut]). Far from being a common and natural assimilation, this variation in the pronunciation of /l/ is not found in many of the world's languages. Thus German *kalt* ('cold') and Italian *caldo* ('warm') are pronounced with clear [l].

In considering the diversity of allophonic adjustment, we should also not forget that languages undergo sound changes, with the consequence that what seems a natural pronunciation to one generation becomes less so to the next. It is fairly clear from the history of the French language, for example, that /l/ had a dark allophone in medieval French. Indeed, the velarization of this allophone was so extreme that it eventually became a [u] vowel (compare the London and South Australian pronunciation mentioned above). We find *u* for earlier *l* in modern French *chevaux* ('horses') (singular *cheval*, earlier plural form *chevals*); and note also *paume* ('palm'), *loyauté* ('loyalty') and *faute* ('fault'). (In all of these instances, the vocalized [u] formed a diphthong with the preceding [a], which has been reduced to a simple [o] vowel in modern French pronunciation.) But with the vocalization and loss of the medieval occurrences of the dark variant, modern French no longer has clear and dark allophones of /l/, and the London English pronunciation of words such as *halt* and *will* does not come easily to French learners.

The range of allophonic variation encountered in natural languages means that it is not easy to predict which sounds can or cannot be allophones of a single phoneme. Some attempts have been made to draw up charts or tables of similar or 'suspicious' sounds. Pike (1947), for example, includes a chart designed as a guide to field workers engaged in transcription and analysis of hitherto unwritten languages. Pike's chart is so complex, however, with circles enclosing sounds judged to be phonetically similar, that it is unlikely to be of much help to any field worker who is not already familiar with the articulatory and auditory character of the sounds referred to. At any rate, there is no mechanical procedure by which one can determine, for any two sounds, whether or not there is at least one language in the world which counts them as variants of a single phoneme. A few general remarks are nevertheless appropriate.

In the first place, it is evident that complementary distribution is not of itself a guarantee that two sounds are allophones of one phoneme. In other words, allophones must show *some* degree of phonetic similarity as well as being in complementary distribution. In some varieties of English, for instance, [h] and [ŋ] are in complementary distribution, since [h] occurs only at the beginning of a syllable (*hat*, *ahead*, *behind*, etc.) whereas [ŋ] is never syllable-initial but always syllable-final or before a consonant (*sing*, *sink*, etc.). (Some speakers of English may have a different patterning, if, for instance, they pronounce *dinghy* with [ŋ] beginning the second syllable.) But even if [h] and [ŋ] are in complementary distribution, they are quite dissimilar in their phonetic nature and it would seem to fly in the face of any sensible description of English to suggest that these two sounds are variants of one phoneme simply because they are in complementary distribution. There are thus limits on the sounds which can be allophones, even though we need to be cautious in giving a universally valid specification of these limits. (See, for instance, Gudschinsky et al. 1970, for a description of a Brazilian Indian language, Maxakalí, in which plosives appear to have vowel allophones.)

Secondly, failure to take account of degrees of phonetic similarity among sounds could lead to patently ridiculous statements. Especially where a language displays general phonetic processes such as nasalization of vowels before nasal consonants, or voicing of plosives between vowels, there will be a number of allophones in complementary distribution with a related set of allophones. Consider, for example, a language in which voiceless [p] [t] [k] occur only word-initially and word-finally, while voiced [b] [d] [g] occur only word-medially. Such a language might have words such as

[pabat] [tadak] [kadap] [pagap] [tabat] [kagak] etc.

Imagine now a computer instructed to scan these words for complementary distributions. The computer would in fact register nine such distributions:

[p] with [b], [p] with [d], [p] with [g];
[t] with [b], [t] with [d], [t] with [g];
and [k] with [b], [k] with [d], [k] with [g].

The correct pairings are of course [p] + [b], [t] + [d] and [k] + [g], but the computer would have no way of recognizing this without some appeal to the kind of phonetic process involved or some insight into the fact that [b] is the voiced counterpart of [p], not of [t] or [k], and so on.

Thirdly, even when the notions of complementary distribution and phonetic similarity are properly combined, there is still room for doubt in some instances about the correct phonemic analysis. Italian, for instance, has three nasal consonant phonemes: /m/ as in *amore* ('love'), *ramo* ('branch'); /n/ as in *anello* ('ring'), *vano* ('futile'), *sano* ('healthy'); and palatal /ɲ/ written *gn* in *agnello* ('lamb'), *ragno* ('spider'), *bagno* ('bath'). Italian also has occurrences of the velar [ŋ] but this sound is found only before velar consonants, written as *n* in e.g. *banca* ('bank'), *lungo* ('long'), *cinque* ('five'). Now although the spelling identifies this [ŋ] as an *n*, it could in fact also be an allophone of /m/ or /ɲ/, as none of the nasal consonants other than [ŋ] ever precedes a velar. Judged by its articulatory position, [ŋ] is actually closer to palatal [ɲ] than it is to [n]. Nevertheless, the solution implied by the standard orthography, namely that

[ŋ] is an allophone of /n/, is widely accepted, even by those whose phonetic interests make them relatively sceptical of phonological analysis (see e.g. Jones 1962, p. 63, on Italian and Spanish [ŋ]). For an instance of alternative solutions in German, see Trim's note (1951) on the fricatives [ç] [x] and [h] in that language: the usual view of German is that [ç] and [χ] are allophones of /χ/, distinct from /h/, but it is also possible to take [χ] and [h] to be allophones of /χ/, distinct from /ç/.

There are thus certain indeterminacies about phonemic analysis. For some linguists, this means that the concept of the phoneme needs refinement, and we turn to some of the issues later in this chapter. For others, as we shall see in chapter 5, the very concept becomes questionable.

4.4 Phonemic norms

If allophones, at least in a large number of cases, are conditioned by their phonetic environment, it seems reasonable to maintain the perspective adopted in section 4.1 and to speak of allophones as variations from a norm. If English /w/ is actually voiceless after voiceless plosives (as it usually is in e.g. *twin*, *quit*), we may say that voiced [w] is the norm but that the normal [w] is 'devoiced' or 'becomes voiceless' under the influence of a preceding voiceless plosive. It thus seems natural to call the phoneme /w/ rather than /w̥/.

It will frequently be the case that one of the allophones of a phoneme readily suggests itself as the normal value in this fashion. The phoneme may then be labelled or transcribed with the symbol representing this normal allophone. In somewhat more technical language, the phonemic symbol should be the symbol of the allophone which is least restricted in its distribution (Pike 1947, p. 88). Two simple examples of the application of this principle are: (1) if the two allophones of a single phoneme are [ŋ] before a velar consonant, and [n] elsewhere, then the phoneme is /n/ rather than /ŋ/; and (2) if the two allophones of a single phoneme are [ã] before a nasal consonant, and [a] elsewhere, then the phoneme is /a/ rather than /ã/. The very use of the term 'elsewhere' of course suggests that the second allophone has the less restricted distribution.

It should, however, be noted that from the perspective of the language in question a phoneme is not necessarily identified with any of its allophones. Moreover, decisions about how to symbolize phonemes are frequently tied up with orthographic issues, not all of which relate directly to phonology. For example, Australian Aboriginal languages in the southern half of the continent usually have no contrast between voiced and voiceless plosives: each plosive phoneme has voiced and voiceless allophones. If these allophones are more or less equally distributed, say voiced allophones word-medially and voiceless allophones word-initially, there may be no particular reason to take either allophone as the norm. Certainly so far as a practical orthography is concerned, it makes little difference whether the spelling employs voiced or voiceless symbols provided it uses one or the other consistently. Indeed, some Australian languages are usually written with voiceless symbols, others with voiced. (The real complications arise where Aborigines who have learned to read and write English introduce into their own language the convention of distinguishing between voiced and voiceless symbols, or where English speakers have transcribed Aboriginal words using both voiced and voiceless symbols on the assumption that there must inevitably be such a distinction. Thus alternative spellings of tribal names such as *Pintupi*, *Bindubi* and *Pindubi*, or *Warlpiri* and *Warlbiri*, continue to compete with each other.)

In some parts of the world, new orthographies have been deliberately designed in ways that conform to an already widely known spelling system. In areas of Latin America where Spanish is the national language, indigenous languages may follow Spanish orthographic conventions even where this is not necessary on phonemic grounds. The phoneme /k/, for instance, may be written as *c* before *a*, *o*, *u*, but as *qu* before *i* and *e*, simply because this follows a Spanish spelling rule with which many readers will already be familiar. Hence, although the selection of a basic allophone or phonemic norm may be important for a phonemic analysis and transcription, orthography is likely to be constrained by other factors.

4.5 Pattern and symmetry

In discussing vowels (section 2.7 above) we noted that systems tend to be symmetrical. Other phonemes may likewise form symmetrical patterns when charted according to their articulatory characteristics. Thus the English plosives (excluding affricates) form a 3 × 2 set, as shown in table 4.5.1(a). In general, languages appear to favour this kind of symmetrical exploitation of contrasts. German, for example, has the same plosive contrasts as English; French and Italian have a similar pattern, except that the voiceless plosives are normally unaspirated and the alveolars tend to be articulated further forward, i.e. as dentals. Some languages distinguish more than just voiced and voiceless plosives, and more than three points of articulation. Some examples are given in table 4.5.1(b)–(d) (again excluding affricates and affricated stops).

Allophones are often similarly patterned. If one voiced stop has a partially devoiced allophone in word-final position, it is highly likely that other voiced stops are subject to the same general phonetic process. Thus in English not only /b/ but also /d/ and /g/ may be partially devoiced at the end of an utterance. If one alveolar consonant is fronted or retracted in certain environments, it is highly likely that other alveolars will behave in the same way. In English, not only /t/ but also /d/ and /n/ are fronted to dental position when immediately preceding a dental fricative. This simply means that allophones tend to be governed by general rules or strategies of pronunciation rather than by idiosyncratic adjustments to individual phonemes.

It is sometimes argued that symmetrical patterning is a target towards which phonological systems keep moving. Certain processes of sound change indeed

Table 4.5.1 Plosive phonemes

(a) ENGLISH

	Bilabial	*Alveolar*	*Velar*
Voiceless (aspirated)	p	t	k
Voiced	b	d	g

(b) KOREAN[a]

	Bilabial	*Alveolar*	*Velar*
Strongly aspirated voiceless	ph	th	kh
Weakly aspirated	p^h	t^h	k^h
Glottalized	p	t	k

(c) HINDI

	Bilabial	*Dental*	*Retroflex*	*Velar*
Voiceless aspirated	p^h	t^h	$ʈ^h$	k^h
Voiceless	p	t	ʈ	k
Voiced aspirated	b^h	d^h	$ɖ^h$	g^h
Voiced	b	d	ɖ	g

(d) ANCIENT GREEK

	Bilabial	*Dental or alveolar*	*Velar*
Voiceless aspirated	p^h	t^h	k^h
Voiceless	p	t	k
Voiced	b	d	g

[a] For the terms used, see Chomsky and Halle (1968, p. 327).

seem to favour symmetry. It seems fairly clear, for example, that in Old English voiced fricatives were not separate phonemes but allophones of the voiceless fricatives: fricative phonemes were voiceless in some contexts, voiced in others. Changes in the language have led to the emergence of separate voiced and voiceless fricative phonemes, namely /f/, /v/, /θ/, /ð/, /s/ and /z/. (Even though the Old English conditioning no longer applies, modern English does still show traces of the earlier pattern, with voiceless fricatives word-final in e.g. *knife*, *half*, *bath*, *south*, *house*, but corresponding voiced fricatives in *knives*, *halves*, *baths*, *southern*, *houses*.) But by the time the voiced fricatives had achieved phonemic status in English, another voiceless fricative /ʃ/ had also arisen, for example by coalescence of /s/ with a following consonant (compare Old English *scip*, *sciell* with modern *ship*, *shell*). This fricative was potentially without a voiced partner, but occurrences of /ʒ/ have in fact been supplied either by new words of French origin (*beige*, *rouge*) or by assimilation of /zj/, as in *measure*, *treasure*, etc. Thus the /ʒ/ has filled what might otherwise have been a 'gap' in the phonemic pattern, as shown in table 4.5.2(a).

Similar arguments to the effect that languages tend to fill 'holes in the pattern' or to maximize symmetrical exploitation of contrasts have been based on various data. Table 4.5.2(b) gives the fricative phonemes of modern German (assuming that [h] is an allophone of /χ/). Each of the voiceless fricatives now

Table 4.5.2 Fricative phonemes

(a) ENGLISH (excluding /h/)

	Labio-dental	*Dental*	*Alveolar*	*Postalveolar*
Voiceless	f	θ	s	ʃ
Voiced	v	ð	z	ʒ

(b) GERMAN

	Labio-dental	*Alveolar*	*Postalveolar*	*Palatal*	*Uvular*
Voiceless	f	s	ʃ	ç	χ
Voiced	v	z	ʒ	j	ʁ

has a voiced counterpart, but from different origins: /v/ results from a change in pronunciation of earlier /w/; /z/ is from earlier /s/; /ʒ/ occurs only in borrowings such as *Journal*, *Manege*; /j/ is the palatal semivowel but is often pronounced with friction so that it virtually becomes a voiced palatal fricative; and /ʁ/ results from the relatively recent adoption of a uvular articulation for earlier /r/. Hence it can be argued that various shifts in pronunciation, some of them ostensibly independent changes to individual consonants, are part of a systemic trend. The classic discussion of this topic is found in Martinet (1955); Fischer-Jørgensen (1975, pp. 44–8) provides a useful overview and additional references.

It is evident, however, that phonemic systems are not always symmetrical. (Indeed, the historical discussion of English and German presupposes that some sound changes destroy rather than create symmetry, otherwise there would never be 'gaps' to be filled.) Voicing contrasts, for example, are not always exploited as systematically as one might expect from simple assumptions about symmetry and economy. In Dutch there is no voiced velar stop: orthographic *g* represents a uvular fricative, and [g] occurs only as an allophone of /k/, whereas /p/, /b/, /t/ and /d/ are separate phonemes. In standard Arabic there are voiced and voiceless sounds in contrast, such as /t/, /d/, /s/, /z/, but no /p/ in contrast with /b/, and no /g/ alongside the /k/.

Moreover, there is always a danger that discussion of phonological symmetry will be more concerned with patterns on paper than with genuine insight into the phonological system. It may well be convenient to represent vowel systems as squared arrays, as in table 4.5.3 (cf. section 2.7 above); but while these diagrams have some merit in displaying the number of vowel contrasts, they have serious drawbacks so far as the nature of the contrasts is concerned. The apparently equivalent systems of Spanish, Russian and Japanese, for example, are rather different in detail. Japanese /u/ is noticeably unrounded, whereas Spanish /u/ and Russian /u/ *are* rounded; Russian /i/ is subject to considerable allophonic conditioning and in many environments is central rather than front, whereas this is not true of Spanish or Japanese; moreover, the effects of such phenomena as stress are quite different among the three languages (with Russian, for example, reducing some unstressed vowels to something like the English indeterminate [ə]); and so on.

Table 4.5.3 Vowel phonemes (squared arrays)

(a) THREE-VOWEL SYSTEM (e.g. Warlpiri, central Australia)

	Front		*Back*
High	i		u
Low		a	

(b) FIVE-VOWEL SYSTEM (e.g. Spanish, Russian, Japanese)

	Front		*Back*
High	i		u
Mid	e		o
Low		a	

(c) SEVEN-VOWEL SYSTEM (e.g. Italian)

	Front		*Back*
High	i		u
High mid	e		o
Low mid	ɛ		ɔ
Low		a	

(d) EIGHT-VOWEL SYSTEM (e.g. Turkish)

	Front unrounded	*Front rounded*	*Nonfront unrounded*	*Nonfront rounded*
High	i	y	ɨ	u
Nonhigh	e	ø	a	o

Even more seriously, a neatly arranged diagram does not necessarily reflect neatly arranged pronunciation. The plosive, fricative and nasal consonant phonemes of French, for instance, can be set out as in table 4.5.4. The arrangement shows that there are three distinctive points of articulation for each kind of consonant; it does not show that each of the three columns represents an identical point of articulation. /f/ and /v/ are labio-dental whereas /p/, /b/ and /m/ are bilabial, and /k/ and /g/ are velar whereas /ʃ/, /ʒ/ and /ɲ/ are palatal. The heading 'Back' above the third column is legitimate in so far as it indicates that all the consonants in this column are articulated further back than alveolar, but certainly not accurate as a precise articulatory label.

Table 4.5.4 Plosive, fricative and nasal consonants of French

	Labial	*Dental/alveolar*	*Back*
Voiceless plosives	p	t	k
Voiced plosives	b	d	g
Voiceless fricatives	f	s	ʃ
Voiced fricatives	v	z	ʒ
Nasals	m	n	ɲ

For further discussion of labels of this kind, see the treatment of phonological and phonetic features in chapter 10 below. In addition to Martinet's work mentioned above, Trubetzkoy (1939, esp. ch. 4) and Hockett (1955, esp. pp. 82–126) give detailed comparative discussion of phonological systems. Section 4.12 below comments further on surveys of sounds across the world's languages. See also other discussion of vowel systems in Liljencrants and Lindblom (1972), Lindau (1978) and section 2.7 above.

4.6 Phonological reality

Enough has been said already to demonstrate that phonological organization is more than a matter of how sounds are articulated. The judgement that English [tʃ] is an affricate but [ts] is not is not based simply on observation and measurement of the way in which these sounds are pronounced or perceived, but requires reference to English sequential patterning and to the phonological system within which these sounds function (section 3.12 above). Likewise, the recognition that clear and dark variants of /l/ are allophones of the one phoneme in English but [h] and [ŋ] are not (section 4.3 above) also depends on more than just articulatory and acoustic observation.

Nevertheless, especially in the English-speaking world, where empiricism and pragmatism are powerful philosophical currents, some linguists have remained suspicious of ascribing any kind of reality to phonological analyses. Some of the scepticism is framed in terms that suggest that articulatory and acoustic phonetics deal with the 'real' or 'objective' nature of speech, while phonology is 'speculative' or 'metaphysical' or 'merely concerned with orthography'. An example is chapter 29 of Jones (1962), where a 'physical' view of the phoneme is defended against a 'superphysical' view. But it is worth noting that few if any of us are totally consistent on such issues. Daniel Jones announces his scepticism about phonemic theory but none the less resorts to an appeal to native speakers' 'feelings' in the case of the Italian velar nasal (1962, p. 63; cf. section 4.3 above).

It is now generally agreed that the classic attempt to produce phonological descriptions that would make no reference to the meanings of words, let alone to native speakers' intuitions or insights, is indeed inconsistent. Z. S. Harris's *Methods in structural linguistics* (1951) represents the claim that it is possible to discover phonemes purely by examining the distribution of phonetic segments: 'The present survey is thus explicitly limited to questions of distribution' (p. 5). But Harris's analysis in fact assumes the investigator's ability to judge whether two utterances in a language are intended to be different words or whether they count as alternative ways of saying the same word. It can be argued that Harris's and others' efforts to define 'objective' analytical procedures constantly presuppose access to native speakers' intuitions into their own language (Chomsky 1964).

Our own view is that it is valid to appeal to the reality of a phonological analysis provided that it is supported by empirical evidence. Empirical evidence can be gathered not only by instrumental means (for example in the spectrographic analysis of sound waves or in the electromyographic analysis of speech organs) but also by the observation of speakers' intuitions. Of course 'intuitions' does not refer here to idiosyncratic or speculative comments about language but rather to what underlies speakers' abilities to count the number of syllables in a word, to say whether two words are pronounced identically or differently, to select rhyming words, and so on. In this sense, the phonological system of a language is open to empirical validation, inasmuch as speakers demonstrate, implicitly or explicitly, their awareness of phonemic differences in their own language. Sapir was particularly intrigued by evidence of this general kind, as for example in the case of a speaker of southern Paiute who pronounced a word as [pa:βah] but then separated it into the two syllables [pa: pah]. This evidence that the native speaker counts [β] as a realization of /p/ (provided that it is not merely an idiosyncratic response on the part of an individual) is just as empirical as the evidence of spectrography or radiography (Sapir 1933).

Admittedly the status of such evidence may be complicated by various factors, including the existence of conventionalized spelling systems, traditions of grammatical terminology and so on. Thus when English speakers say that English has five vowels they are referring to the five letters A E I O U and not to the phonological system. On another level, however, the same English speakers operate with more than five vowels when they construct or assess rhyming verse (in which case they respond to phonemic contrasts rather than spellings). Similarly, English speakers may claim that the words *cent*, *sent* and *scent* are different because they have different spellings and meanings. But they will agree that they are pronounced identically – or, putting it in an empirical context, they will be unable to distinguish the words when given only the pronunciation and not the spelling or meaning.

We have already referred to such evidence in connection with interpretations and phonemic analysis (sections 3.14 and 4.2 above) and we stress that the variety of available evidence points not only to different levels of analysis but also to interaction and integration among these levels. Even without access to instrumental findings about the articulation and acoustics of speech, speakers are aware of interrelationships in their language, say between words which differ in meaning but not in spelling and pronunciation (*football matches*, *box of matches*) or between different spellings of the 'same' word (*Catherine*, *Katherine*; *judgment*, *judgement*) and so on. One could imagine a language in which all of this was maximally simple: each phoneme would have a single allophone, with minimal variation in articulation and acoustic properties, the spelling would have a perfectly consistent one-to-one mapping of visual symbols on to phonemes or syllables, there would be no synonymy or homonymy, and so on. In practice, although some languages are simpler or more consistent in certain respects than others, maximal simplicity seems to be so remote from the truth that it is artificial.

4.7 Units and boundaries

Many phonological processes apply within certain domains. For example, the lengthening of English /n/ before a voiced stop or fricative (as observed in words such as *sand*, *bend*, etc.) does not apply where the nasal and obstruent belong to different syllables. There is no lengthening of the /n/ in *undo* or *indecent* compared with *until* or *intelligent*. We may say that this lengthening is 'intrasyllabic', i.e. it applies within a syllable. On the other hand, some English processes clearly have a larger domain. The assimilation of /n/ to the point of articulation of a following consonant is in no sense blocked by a syllable or word boundary, and the /n/ in *ten boys* or *ten miles*, for instance, may often be pronounced as [m].

In fact a proper account of phonology, including intonation, stress and assimilatory processes as well as phonemic contrasts, requires reference to units at various levels. Many linguists recognize an ascending hierarchy of units such as: syllable, phonological word, tone group, breath group, etc. (See section 3.1 above and remarks on the organization of intonation in section 9.8 below.) Note that the boundaries of these units do not necessarily coincide with grammatical boundaries. It can be argued, for instance, that an English article plus a noun form a single phonological word, even though there are two distinct grammatical elements written as two words. The article is normally unstressed and is phonologically indistinguishable from a prefix: compare *a head*, *a way* with *ahead*, *away*. Actually the history of certain English words makes it quite clear that the boundary between article and noun is not a strong one: *adder* (snake) and *apron* are derived from earlier forms *nadder* and *napron*, by a process in which *a nadder* and *a napron* were taken to be *an adder* and *an apron*. Common reduced forms such as *I've*, *he's*, *she'll*, *don't* also demonstrate that grammatical and phonological units need not coincide: each of these forms is a single syllable but two grammatical elements (MORPHEMES).

This is not to say that phonological and grammatical units never coincide. There are certain languages (and certain phenomena within particular languages) in which grammatical units have special relevance to phonology. A simple instance is the Javanese glottal stop occurring as an allophone of /k/ in morpheme-final position. (This differs from Indonesian, in which /k/ has the glottal stop allophone in *word*-final position, as described in section 4.3 above.) Note the following Javanese words, where hyphens have been added to the normal spelling to show the morphemic composition:

anak	[anaʔ]	child
anak-e	[anaʔe]	the child
mangan-ake	[maŋanake]	cause to eat
temok-ake	[təmɒʔake]	cause to meet

To interpret /k/ correctly as [k] or [ʔ], one must know whether it is at the end of a morpheme.

In Turkish, a process of vowel harmony extends through the word. In general outline, it is the vowel of the first syllable that is distinctive, and the vowels of subsequent syllables are constrained within the rules of the language. As a consequence, suffixes have different phonemic shapes, depending on the vowels of the roots to which they are attached. Some suffixes, such as plural, have a front *e* vowel if preceded by a front vowel; otherwise the suffix has the *a* vowel. Other suffixes, such as the genitive, have four different vowels, again depending on the nature of the preceding vowel in the root: these suffixes have front unrounded *i* after *i* or *e*, back rounded *u* after *u* or *o*, and so on. Table 4.7.1 gives some examples in standard Turkish spelling, in which front rounded vowels are shown by a diaresis above the *u* and *o* and a high central or back unrounded vowel is represented by an undotted *i*.

Turkish actually has two vowel systems, effective at different points of the structure. In the first syllable of a root, any of a full set of eight vowels can occur. In subsequent syllables (including suffixes, of course) there is systemically only a two-way choice between a relatively low vowel (which has two variants according to context) and a relatively high vowel (which has four variants according to context). Table 4.7.2 shows these systems. One consequence of this is that Turkish suffixes – units recognized in the grammar of the language – are not fully specified for vowel quality but depend on the root to which they are affixed.

Table 4.7.1 Examples of vowel harmony in Turkish

Root	*Meaning*	*Root + plural*	*Root + genitive*
kedi	cat	kediler	kedinin
ev	house	evler	evin
kız	daughter	kızlar	kızın
adam	man	adamlar	adamın
gün	day	günler	günün
göz	eye	gözler	gözün
ulus	nation	uluslar	ulusun
kol	arm	kollar	kolun

Table 4.7.2 Turkish vowels

(a) FULL SYSTEM (in first syllable of a root)

	Front		*Central/back*	
	Unrounded	*Rounded*	*Unrounded*	*Rounded*
High	i	ü	ɪ	u
Low	e	ö	a	o

(b) SUBSYSTEM (in noninitial syllables, including suffixes)

High	I
Low	A

I is realized as /i/, /y/, /ɨ/ or /u/ according to harmony.
A is realized as /e/ or /a/ according to harmony.

Thus phonological description must sometimes take account of grammatical units, such as morpheme or suffix; and grammatical description may sometimes need to recognize the phonological properties of grammatical units. (For more general remarks on the interaction between phonology and grammar, see section 4.10 below.)

4.8 Invariance and overlap

A rigid model of phonemic organization can give the impression that every phoneme has certain invariant characteristics. Thus it might be supposed that English /p/, despite some allophonic variation in the degree of aspiration and the nature of the plosive release, will be invariably bilabial, voiceless and plosive in character. While this may be reasonable for the specific case of English /p/, it is simplistic to assume that comparable invariant features can be specified for every phoneme in all languages.

In the first place, it is sometimes extremely difficult to specify precisely what features are common to all allophones of a phoneme. English /r/ may have allophones ranging from a voiced tap or flap, to a voiced (frictionless) approximant, to a voiceless fricative. If there are common characteristics shared by all of these allophones, they are more easily defined in negative terms (nonlateral, non-nasal, nonvelar, etc.) than in precise phonetic terms. Indeed, this is one reason why phonological description frequently resorts to terminology which is language-specific, if not *ad hoc* (sections 4.5 above and 10.7 below).

In the second place, linguistic distinctions are relative rather than absolute. For the sake of simple illustration we take an artificial example, which is nevertheless based on the kind of phenomena encountered in a number of natural languages. The following words reveal four phonetic vowels but only three contrasts in any particular environment:

[tip]	[tɪk]	[pit]	[pɪk]
[tɪp]	[tek]	[kɪp]	[kek]
[tep]	[tɛk]	[ket]	[pɛk]

Minimal pairs demonstrating the three vowel phonemes are contained in the first two columns. The reasonable explanation of this language is to say that there are three vowel phonemes, each of which is lowered before [k], i.e.

/i/ [ɪ] before [k]
[i] elsewhere
/ɪ/ [e] before [k]
[ɪ] elsewhere
/e/ [ɛ] before [k]
[e] elsewhere

The phonemes overlap with each other, in that one allophone of /i/ is identical with one allophone of /ɪ/, and one allophone of /ɪ/ is identical with one allophone of /e/. But distinctions are maintained, because the contrast is one of *relative* vowel height in the relevant context: whether before a [k] or not, /e/ is always lower than /ɪ/, and /ɪ/ always lower than /i/. Vowel systems often show shifted contrasts in this manner, and data of this kind are attested for various languages (see Jones 1962, ch. 19, for examples from French and Russian, and Stokes 1981, esp. pp. 149ff., and Waters 1979, pp. 69ff., for the Australian languages Anindilyakwa and Djinang). Examples of this kind obviously defy any attempt to specify the *absolute* values of each phoneme.

In the third place, contrasts are not always localized strictly within one segment. The contrast between English /t/ and /d/, for example, is often more a matter of the length of the preceding segment than of the nature of the plosives themselves. Compare pairs like *seat* and *seed*, or *bent* and *bend*, and note that in certain circumstances (say over a bad telephone line) the length of the preceding vowel or nasal consonant is likely to be a more crucial factor than the quality of the plosive ('I said seeeed, not seat'). Phonological systems do not appear to be constrained by a principle that distinctions must be firmly anchored within segmental boundaries, and there are many other examples which may raise doubts about too narrow a concept of the segment as a basic unit (section 3.1 above). In Javanese, for example, the distinction between voiced and voiceless plosives often seems to be signalled by the nature of voicing in the following vowel (breathy voice after a voiced plosive). And in many varieties of German (especially in the north of the country) the presence of a final *r* is indicated by the quality of a preceding vowel (mimicked by other Germans as *bessa* instead of *besser*, *guta* instead of *guter*, etc.).

The historical developments which many languages have undergone further demonstrate the relativity of phonemic distinctions. There is ample evidence of quite radical shifts in the nature of these contrasts. A system of long and short vowel contrasts may at some later stage of the language become a system of pure and diphthongal vowel distinctions. Or if consonants are dropped or elided, vowel allophones that were conditioned by the lost consonants may become contrastive vowel phonemes; and so on. Many tonal languages, for instance, can be traced back to an earlier stage at which pitch was a redundant feature associated with certain adjacent consonants (Hyman 1975, pp. 228–9, and section 9.4 below).

One of the reasons for the diversity of modern English pronunciation is that the vowel system has undergone major shifts over the last few hundred years, with different consequences in various regions. Some five hundred to six hundred years ago the English vowels of e.g. *time* and *tame* were approximately [i] and [a]. (Compare the values of the letters *i* and *a* which persist in other European languages such as French and German.) In modern English pronunciation these vowels are commonly diphthongs, although the extent and nature of diphthongization vary considerably. At the same time, the loss of final /r/ in south-eastern England has led to modification of preceding vowels, e.g.

here	earlier /hi:r/	modern /hɪə/	(cf. *he* /hi:/)
hire	earlier /hair/	modern /haɪə/	(cf. *high* /hai/).

The forms given here as 'earlier' are in fact maintained in some parts of the English-speaking world (notably south-western Britain and much of North America). But areas where English has spread from Britain subsequent to the elision of final /r/ or where closer connections have been maintained with British English (such as Australia and New Zealand) show the same kinds of contrast as modern south-eastern British English.

These examples show that different historical stages and regional varieties of a language may have different phonological organization, and they underline the point that a phonemic system is a network of relative contrasts. They do not, however, rule out the possibility of unambiguous phonemic analysis for any language taken as a particular regional version at a particular point of time. In the following section, however, we turn to phenomena that can create serious ambiguity in the analysis itself.

4.9 Biuniqueness and neutralization

A phonemic description is said to be BIUNIQUE if phonemes and allophones are unambiguously mapped on to each other. The analysis of the three vowel phonemes in the artificial data in section 4.8 above is biunique, despite some overlap, because environments can be clearly specified: [ɪ] before [k] is unambiguously an occurrence of /i/ whereas [ɪ] before consonants other than [k] is equally clearly an allophone of /ɪ/. Admittedly, where there is linear realignment of a contrast, reanalysis may be necessary in order to preserve biuniqueness. Suppose that English pronunciation actually changed to the point where the only distinction between final /t/ and /d/ (and other voiceless and voiced sounds) was in the length of the preceding segment, i.e.

send pronounced as [sɛn:t], *seed* as [si:t];
sent pronounced as [sɛnt], *seat* as [sit].

Now it is difficult to contrive a statement of phonemes and allophones to the effect that /t/ is [t] but that /d/ also has an identical allophone [t] provided that the preceding segment is lengthened. Moreover, a simple statement that both /t/ and /d/ are sometimes indistinguishable as [t] would violate the principle of biuniqueness. It would therefore be more realistic to recognize that words no longer end in /d/ and that the language now has new phonemic contrasts such as /n/ versus /n:/.

For any particular system, then, biuniqueness is a requirement that phonemes and allophones can be unambiguously assigned to each other. A problem in this connection is that contrastive systems are often unequally exploited. This means, for example, that two phonemes may be distinguished in some structures but not in others. Following Trubetzkoy (1939) we may say that some phonemic oppositions are suspended or NEUTRALIZED under certain conditions. Trubetzkoy distinguishes three kinds of neutralization and we give examples of each.

First, a language has a certain contrast but only one of the relevant phonemes occurs under neutralization. Suppose a language has a contrast of voiced and voiceless plosives in word-initial and word-medial positions, but only voiceless plosives occur word-finally. Since the word-final plosives are not in contrast with voiced plosives, the contrast of voicing is inoperative or neutralized word-finally. This pattern of neutralization is found in a number of languages, including Dutch, German and Russian. In Dutch, /t/ and /d/ are in contrast, e.g. in *toen* ('then'), *doen* ('to do'), *teken* ('sign'), *deken* ('blanket'). Although written forms show both final *t* and *d*, there is no such thing as a final voiced plosive in pronunciation. Thus *bond* ('association') and *bont* ('fur') are pronounced identically, with final [t], as are *pond* ('pound') and *pont* ('ferry'). Comparable illustration of the same pattern of pronunciation can be found in German and Russian, although the details of how the neutralization applies and how it intersects with assimilatory processes of voicing and devoicing vary from language to language. (In particular the concept of 'word-final' neutralization needs refinement, since the neutralization may apply, for example, at the end of the first element of a compound as well, as in Dutch *bondgenoot* 'ally, confederate'.)

Secondly, neutralization may be represented by some kind of variation or alternation among the otherwise contrasting phonemes. In Indonesian, for example, there are four nasal consonant phonemes (bilabial /m/, alveolar /n/, palatal /ɲ/ and velar /ŋ/); but sequences of nasal plus other consonant are homorganic, that is, the nasal and following consonant are at the same point of articulation. Thus we find clusters such as /mb/ and /nd/, but not /md/ or /nb/. (Borrowing from other languages has brought some exceptions to this pattern, but we ignore these for the sake of illustration.) This means that there is no contrast of nasal consonants preceding a plosive. It would be possible to represent the preconsonantal nasal with a single symbol (say *n* or *N*): the value of *N* would be entirely predictable from the point of articulation of the following plosive. Other languages in which nasal consonant clusters are similarly homorganic include Japanese and Spanish.

Alternatively, neutralization may be represented by free variation (section 4.11 below) of the phonemes in question. Some varieties of English have a contrast between /aʊ/ and /aʊə/ in e.g.

cow	[kaʊ]	cower	[kaʊə]
bow	[baʊ]	bower	[baʊə]

This contrast is neutralized before /r/ (and often also before /l/), where there may be indeterminate variation between the diphthong and triphthong. For example, Australian students beginning to learn to transcribe English are often uncertain whether they say the place name *Cowra* as /kaʊrə/ or /kaʊərə/. Similar indecision usually affects words such as *dowry*, *cowering*, *towel* and *owl*.

Thirdly, neutralization may be represented by a sound which is distinct from both of the otherwise contrasting phonemes. One of the most common instances of this kind of neutralization is where vowel contrasts are reduced under certain conditions, say before certain consonants or in unstressed syllables. The English tendency to reduce all vowels to the so-called 'indeterminate' [ə] is one illustration of the principle. Compare the capitalized vowels in e.g.

legAlity [æ]	legAl [ə]
irOnic [ɒ]	irOny [ə]
torrEntial [ɛ]	torrEnt [ə]

In varieties of English such as RP, [ə] never occurs as the vowel of a fully stressed syllable (other than as the offglide of centring diphthongs). It can therefore be seen as representing neutralization of the usual range of vowel contrasts.

It should be noted that neutralization sometimes creates alternate forms of a morpheme: English *torrent* has two different phonemic forms depending on whether it is unsuffixed or carrying the suffix *-ial*. In this case, the phenomenon may be described as morphophonemic (section 4.10 below).

In terms of a phonological analysis, there are three ways of treating neutralizations. The first is to insist that sounds representing neutralizations must be treated as allophones of a phoneme – which means in effect not recognizing neutralization. Thus if a language has both voiced and voiceless plosives but only voiceless plosives in word-final position, the word-final voiceless plosives are simply taken at face value. It must then be said of this language that voiced plosives do not occur word-finally. While this reflects a phonetic truth and may seem perfectly obvious, some cases will require an arbitrary choice. For instance, where the distinction between /aʊ/ and /aʊə/ is neutralized before /r/ in English, it is not clear by what criterion one can insist that the vowel is phonemically one or other of the two alternatives.

A second possibility, which avoids this arbitrariness, is a strategy proposed by Trubetzkoy himself, namely that of recognizing an ARCHIPHONEME. Thus English [ə] might be judged to be an archiphoneme representing the neutralization of vowel contrasts exhibited in stressed syllables: it is not identified with any of the other vowel phonemes but represents the suspension of the relevant contrasts. In this tradition of analysis, archiphonemes are often indicated by capital letters to show their special status. Applying this convention to a language in which the voicing opposition is neutralized word-finally, we might write final plosives as /P/, /T/ and /K/. We have also mentioned the possible use of /N/ for a nasal consonant that takes the point of articulation of the following consonant. An archiphoneme is in effect an underspecified segment. Thus /N/ stands for 'nasal consonant' without point of articulation features, /P/ for 'bilabial plosive' without specification of voicing, and so on.

A third possibility is to forgo biuniqueness. If we do this, we seem to introduce ambiguity into the analysis. If we suggest, for instance, that English [ə] is indeed an allophone of *any* other vowel, then we may have no way of determining, for any particular occurrence of [ə], to which of the vowel phonemes it is to be assigned. But there are often related forms which do provide a means of making a choice. The very fact that the form *torrent* (with [ə]) is related to the form *torrential* (with [ɛ]) provides a reason for allocating [ə] to the phoneme /ɛ/ in this instance; while a comparison of *irony* and *ironic* allows us to say that the [ə] in *irony* 'belongs to' the phoneme /ɒ/. Now there are indeed multiple

sources or origins for [ə], and each case will be decided by related forms. From a strictly phonemic perspective, the analysis is ambiguous, for there is nothing in the phonological context that tells us that one [ə] belongs to /ɛ/, another to /ɒ/, and so on. But there is no indeterminacy once the appeal to grammatical or semantic relationships is allowed. To return to the example of word-final neutralization of voicing in languages such as Dutch and German, we can distinguish between voiced and voiceless plosives, even though they are pronounced identically. Here too there are related forms to appeal to, e.g.

pond [pɒnt] pound	*ponden* [pɒndən] pounds
pont [pɒnt] ferry	*ponten* [pɒntən] ferries
bond [bɒnt] association	*bonden* [bɒndən] associations
bont [bɒnt] fur	*bonten* [bɒntən] furs.

It is important to note that *pond* and *pont* are not distinct in pronunciation – but once we know the meaning, or specifically affixed forms such as the plural, then we can relate [t] to either /t/ or /d/.

It is interesting to measure biuniqueness in phonemic analysis against the orthographic practice of written languages. Generally speaking, a spelling system that matches or reflects a biunique phonemic analysis is an attractive one. It is the kind of spelling system that is commonly but misleadingly termed 'phonetic': there will be no orthographic ambiguities, so that any letter or symbol will have a unique value (i.e. pronunciation) and any sound will have a unique orthographic representation (letter or symbol). Of course, the pronunciation of some letters will be relative to the environment in which they stand, because some phonemes have various allophones, but the correct pronunciation will be governed by the 'allophonic rules' of the language. Thus it may be necessary to know that word-final *k* is pronounced as a glottal stop in Indonesian or that *l* is dark before a consonant in English, but these are matters of unambiguous rule. Native speakers who have already learned to speak their own language will not need to be instructed in what they take to be the normal way of pronouncing phonemes.

There is none the less a case *against* biunique spelling systems. It can be argued, for example, that a spelling system ought to distinguish homonyms (*knight* and *night*, *right*, *rite* and *write*, etc.) or that a conservative and even difficult spelling system may be justified as a common orthography for speakers of different dialects. Furthermore, there are few if any current orthographic systems that are truly biunique. Even those often praised for their consistency and simplicity, such as Dutch, Italian and Indonesian, have some ambiguities. For example, Dutch *nog* and *noch* are pronounced identically, Italian *e* represents both /e/ and /ɛ/, and Indonesian *e* represents both /e/ and /ə/. The fact remains, however, that most spelling systems approach phonemic biuniqueness much more closely than the notoriously conservative orthographies of English and French.

Few orthographies have special letters corresponding to the archiphonemes of phonological analysis. One possible case is the apparently redundant use of special letters in Ancient Greek to represent the sequences /ps/ and /ks/. Ancient

Greek actually had a contrast of three kinds of plosive, namely voiceless aspirated, voiceless and voiced (table 4.5.1(d) above). /pʰ/, /p/ and /b/ were represented by the letters known as *phi* φ, *pi* π and *beta* β, and /kʰ/, /k/ and /g/ by the letters known as *chi* χ, *kappa* κ and *gamma* γ. (We restrict our attention here to the labial and velar instances, and should also note that Modern Greek does not preserve this three-way distinction, some of the plosives of Ancient Greek now being pronounced as fricatives.) The contrasts of aspiration and voicing were, however, neutralized before /s/. Numerous consequences of this pattern of neutralization can be observed in Ancient Greek. In verb forms, for example, the verb root is followed by /s/ in future forms, as in /lu-/, verb root meaning 'undo' or 'loose', /luo:/ 'I undo', /luso:/ 'I will undo'. Where the verb root happens to end in a plosive, the neutralization will be evident before the /s/ in future forms:

/grapʰo:/ I write	/grapso:/ I will write
/blepo:/ I see	/blepso:/ I will see
/tʰlibo:/ I press	/tʰlipso:/ I will press
/arkʰo:/ I rule	/arkso:/ I will rule
/dio:ko:/ I chase	/dio:kso:/ I will chase
/anoigo:/ I open	/anoikso:/ I will open.

The neutralization means of course that one cannot tell from a future form what the root is – whether /k/ before /s/ corresponds to a root-final /k/, /kʰ/ or /g/. What is noteworthy for our purposes is that the Greek spelling system uses single letters for the sequence of a neutralized plosive and following /s/, namely *xi* ξ for /ks/ and *psi* ψ for /ps/. These special letters can be taken as orthographic signals of the neutralization (cf. Allen 1987, pp. 59–60). Traces of the spelling conventions can be found in English. The flower *phlox* takes its name from the Greek for 'flame', with the letter *x* representing /ks/, where the final /s/ is a suffix; the root actually ends in /g/ (when not affected by a following /s/) as seen in words such as *phlogistic* and *phlogiston*. The occasional use in German of spellings such as word-final *dt* (*Stadt*, *Brandt*) is also a minor instance of special orthographic recognition of a loss of contrast. But in most spelling systems there are few if any special devices to represent neutralization.

It seems rather more common that orthographic practice reponds to related forms. Thus in Dutch, German and Russian, the orthography does distinguish between final voiced and voiceless plosives, depending on how the consonant is pronounced in nonfinal position. Dutch *pond* is written with a *d*, *pont* with a *t*, even though both end in [t] and are therefore indistinguishable in pronunciation: the spelling is justified by appeal to other forms such as the plurals *ponden* (pronounced with [d]) and *ponten* (pronounced with [t]). Likewise, English [ə] is written with various vowels, often depending on a related form: thus we write *e* in *torrent* (compare *torrential*), *a* in *legal* (compare *legality*), and so on. Nevertheless, many users of English are evidently not always sensitive to related forms. Common spelling mistakes include errors like *grammer*, even though a knowledge of the pronunciation and spelling of *grammatical* would suggest the spelling *grammar*. In some instances users may simply not

know the related forms, such as the word *sentential* (justifying the use of *e* rather than *a* in the second syllable of *sentence*). And it must also be recognized that in many cases in English, there is no related form with a full vowel: spellings such as *o* in *button*, *a* in *defendant* and *e* in *apparent* cannot be justified by appeal to other forms.

A brief but useful explanation of neutralization, based on Trubetzkoy's exposition, can be found in Sommerstein (1977, pp. 49–53); see also section 11.6 below.

4.10 Morphophonemic alternations

If a morpheme has two or more phonemic shapes, the different forms are sometimes referred to as ALLOMORPHS (compare the term 'allophone', section 4.3 above). Allomorphs are not necessarily closely similar to each other. In Dutch, for example, the plural suffix is

-en in e.g. *ponden* ('pounds'), *bonen* ('beans')
-eren in e.g. *kinderen* ('children'), *eieren* ('eggs')
-s in e.g. *tafels* ('tables'), *zoons* ('sons').

Even more remote from each other are English forms such as *go* and *went*, where, arguably, *went* (or *wen-*) can be regarded as the allomorph of the verb that occurs in the past tense.

Some allomorphs, however, belong within a general pattern of phonemic alternation. In this case the allomorphs may be said to be in MORPHOPHONEMIC ALTERNATION with each other. We have already met some examples earlier in this chapter, such as the neutralization of final voiced and voiceless plosives in languages such as Dutch and German. In German, words such as *Hunde* ('dogs') and *Hände* ('hands') contain /d/; but the singular forms *Hund* and *Hand*, although written with *d*, are pronounced with /t/. Forms such as *Bund* ('federation') and *bunt* ('colourful') are therefore indistinguishable in pronunciation: it is only suffixed forms such as *Bundes* and *buntes* that show an opposition between /d/ and /t/. Not only are /t/ and /d/ phonologically close, differing only in voicing, but the pattern is a highly regular one: any final /d/ will be devoiced but will be recoverable from related forms in which the /d/ is not final. Moreover, the pattern is not just an alternation of /t/ and /d/ but extends to all voicing contrasts in German. Thus *Laub* ('foliage') has final /p/, but the /b/ is recoverable from e.g. *laubig* ('leafy').

In English, voicing contrasts are also neutralized, giving rise to morphophonemic alternations, although under different conditions from German. English /s/ and /z/ are separate phonemes (*seal/zeal*, *fuss/fuzz*) but the plural suffix is /s/ in words such as *maps*, *cats*, *socks*, and /z/ in words such as *tubs*, *lids*, *dogs*, even though the conventional spelling does not show the difference in pronunciation. This is again part of a wider pattern, applying also, for

instance, to the possessive suffix (as in *the cat's food* and *the dog's food*) as well as to the past suffix (*rubbed*, *sagged* ending in /d/, *ripped*, *sacked* ending in /t/).

Contrasts among English nasal consonants are likewise neutralized under certain circumstances. Notice, for example, that /m/ and /n/ are separate phonemes (*meat/neat*, *sum/sun*) but the prefix in words such as *improbable*, *imbalance*, *indecent*, *insolvent* ends in /m/ or /n/ depending on the following consonant. (Here English spelling does show the difference in pronunciation, whereas written *-s* and *-ed* stand for alternative phonemic forms.)

English also has a number of vowel alternations. The following examples show four pairs of alternating vowels: in each case the forms on the right show a different vowel from the forms on the left, in the syllable immediately preceding the suffix *-ic*:

state, mania	static, manic
esthete, academe	esthetic, academic
analyse, type	analytic, typic(al)
cone, microscope	conic, microscopic.

Phonemically, each pair of vowels is distinct, as shown by minimal pairs such as

mate / mat	main / man	fate / fat
seat / set	dean / den	steam / stem
type / tip	sight / sit	lime / limb
own / on	coat / cot	toast / tossed.

Despite such instances of contrasts, the occurrence of one vowel rather than the other is often predictable from the grammatical context. Thus the same vowel that occurs before *-ic* also occurs under other conditions: for instance, the alternation evident in *esthete* and *esthetic* is also seen in e.g.

obscene	obscenity
convene	convention
keep	kept.

This predictability is often not thought of as an instance of neutralization, for neutralization in the classic sense is peculiar to some specific phonological environment (such as word-final position, or preceding a consonant). In this case, while we can predict the change of vowel when /t/ is suffixed, in forms such as

keep, kept sleep, slept weep, wept

it is not true that the distinction between the two vowels is neutralized before /pt/ (or comparable consonant sequences). Note, for instance forms such as *heaped* and *reaped* and the minimal pair *steeped/stepped*. Indeed, even particular suffixes do not guarantee that the alternation will apply. Thus *-ic* is preceded by /ɒ/ in e.g. *conic* and *tonic*, but the word *rhotic* is often pronounced with /oʊ/, possibly because it is not perceived to contain the suffix *-ic*; the words *phonemic* and *morphemic* are often pronounced with the same vowel

as in *phoneme* and *morpheme*, rather than rhyming with *endemic*; and some Australians pronounce *basic* with the same vowel as *base*, others rhyme it with *classic*.

Phonological analysis in narrowly phonemic terms has often relegated these apparent violations of phonemic consistency to a special category of description intermediate between phonology and morphology: hence the blended term MORPHOPHONEMICS, replacing earlier MORPHONOLOGY or MORPHOPHONOLOGY (Martinet 1965). The allomorphs or variant forms of specific morphemes may then be described under this heading. For instance, the English plural suffix has (among others) the allomorphs

/-s/ occurring after roots ending in /p/, /t/, /k/ etc. (as in *cups*, *pots*, etc.)
/-z/ occurring after roots ending in /b/, /d/, /g/ etc. (as in *clubs*, *heads*, etc.).

Or *telephone* may be said to have (among others) two different allomorphs, depending on whether the stem is unsuffixed or carries the suffix *-ic*.

Generalizations about patterns of alternation can be expressed as MORPHOPHONEMIC RULES. Thus there is a morphophonemic rule of devoicing final stops and fricatives in languages such as German and Dutch; and in English the processes of vowel alternation such as we have illustrated above are sometimes covered by general rules of 'tensing' and 'laxing'. The status and validity of such rules became a key issue in the 1960s, as part of a wider debate about the nature of phonological description (chapter 5 below).

Martinet's classic discussion of morphophonemics (1965) includes a brief review of the origins and early uses of the terminology in the writings of Trubetzkoy (1939) and Bloomfield (1933) (see also sections 11.5 and 11.6 below). A helpful summary can be found in Sommerstein (1977, pp. 41–4).

4.11 Free variation

The notion of free variation or free fluctuation is intended to account for random interchangeability in language. Suppose that an English speaker pronounces the initial consonant of *then*, *this*, *there*, etc., as either a dental fricative or a dental plosive and is unaware of the variation or apparently indifferent to the choice. The two sounds can be described as FREE VARIANTS or freely fluctuating allophones of the phoneme. The allophones are ostensibly *un*conditioned by their phonetic environment (section 4.3 above). The usual notation, implying random variation in any environment, is

/ð/ [ð] ~ [d].

Alternatively, if the free variation applied only in word-initial position, we would then show it as follows:

/ð/ [ð] ~ [d] word-initially
[ð] elsewhere.

In practice, a great deal of free variation will be of the kind that is not noticed even by trained phoneticians, let alone by the average speaker or learner of the language, such as minute variations in tongue position or timing that are a natural part of articulatory processes (section 4.1 above).

Other apparent cases of free variation may be as much due to uncertain hearing on the part of the linguist recording the language as to indifference on the part of the speaker. Still other 'free' variants turn out to be associated with specific regions or styles. English speakers may say that it does not matter whether you pronounce the /r/ phoneme as an approximant or as a flap or trill; but in fact there are strong regional and stylistic associations. Scottish speakers favour the flapped or trilled articulation far more than, say, English or Australian speakers do, and use of the flap is the kind of evidence that enables people to identify regional origins. English and Australian speakers may, however, adopt a flap or trill in certain circumstances, including operatic singing or other kinds of highly deliberate or careful speech. Thus it would be quite wrong to suggest that differences in the articulation of /r/ are a matter of free variation in English: it is true that such differences are not functional within the phonological system of contrasts, but they are not randomly disregarded and certainly are communicative in signalling speech styles or regional identity. Allophonic variation that is truly free probably occurs rarely, if at all, unless it is below the threshold of normal perception.

The concept of free variation may also be applied to phonemes themselves. Consider, for example, the possibility of pronouncing the English word *economics* with either /ɛ/ or /iː/ as the first vowel, or *either* with /iː/ or /aɪ/. The vowels are separate phonemes (compare *head*, *heed* and *hide*, or *men*, *mean* and *mine*) and are not interchangeable in most words. Cases of this kind are more likely to constitute genuinely free variation, especially where neutralization is involved: Australian English speakers, for example, may be undecided between /aʊ/ and /aʊə/ in words like *dowry* and *cowering* (section 4.9 above). Once again, regional and stylistic preferences are often involved. In the case of the word *either*, pronunciation with /iː/ is widely regarded as 'American' and with /aɪ/ as 'British', although this is something of an oversimplification, since both pronunciations can be heard in Australia. Some readers may conclude that this merely confirms that Australia is torn between British and North American models. Certainly, speakers are likely to be more conscious of differences among phonemes than of allophonic variation, and they may indeed be torn between competing norms. Readers will be familiar with the phenomenon of a speaker with a shifting pronunciation, say someone who moves from one area to another and seems to have partially and inconsistently changed pronunciation as a result, or someone who seems sometimes to be 'putting on' a different accent. Such phenomena are of considerable social significance – people are often alert to what they perceive as oddities or signs of an 'outsider's' accent, and may be quick to condemn those who 'betray' their native accent, for example. A simple concept of free variation is inadequate to explain the

complexities of speech communities and the norms towards which individuals aspire.

For further remarks on free variation, see Pike (1947, ch. 11), Harris (1951, pp. 29ff.) and Sommerstein (1977, pp. 18–19). The social and regional significance of speech variation is a large subject in its own right: a general introduction to the sociolinguistic study of variation – including summary accounts of particular studies of phonological variables – can be found in Wardhaugh (1986, esp. chs 6 and 7); and Wells (1982) is a thorough survey (in three volumes) of regional diversity in English pronunciation.

4.12 The sounds of the world's languages

If it is possible to list an inventory of phonemes for any language, then it is also possible to look for generalizations across these inventories, by asking questions such as the following:

- What are the most common kinds of phoneme?
- What is the average number of phonemes in a language?
- Are some phonemes found only in some regions of the world?

Questions such as these have been pursued by a number of linguists, often in connection with an interest in 'universals of language' and more recently in the context of compiling databases of 'phonological segment inventories' such as the one created at the University of California at Los Angeles, known as the UCLA Phonological Segment Inventory Database, or UPSID for short (see the introduction to Maddieson 1984).

Such questions are nevertheless not as easy to answer as one might hope. In the first place, we have to decide what we mean by 'all languages'. There are many languages which we know were once spoken but are no longer in living use. While we have no particular reason to believe that any of these extinct languages was radically different in its phonology from modern languages, it would certainly be unwise to generalize too confidently. In fact, not even all the living languages of the world have been analysed in sufficient depth to allow us to say what their phonemes are, making it all the more necessary to be cautious.

In the second place, statements about what kinds of phonemes all or most languages have – or don't have – may overlook the difficulties of determining what counts as one language. To take a simple illustration from English, most phonologists would say that English has three nasal consonant phonemes: the /m/ of *sum* or *ram*, the /n/ of *sun* or *ran*, and the /ŋ/ of *sung* or *rang*. If we were trying to make some generalizations about how many and what kind of nasal phonemes languages have, we would thus count English as one of the languages which has three nasal consonants. But in fact, as we noted in section 4.3 above, there are regional varieties of English, in the Midlands and north

of England, in which there is no velar nasal phoneme: words like *sung* and *rang* are pronounced with a final /g/ following the velar nasal consonant, and the nasal consonant is therefore an allophone of /n/ conditioned by a following velar consonant. It is clear that any generalizations about how many languages have three nasal consonant phonemes, and how many have two, will be affected by whether we count English as a single language and ignore its regional variation, or whether we begin to recognize regional varieties as (potentially) different phonological systems, and therefore as different languages. There are many examples of this kind, especially in parts of the world where what is commonly referred to as a single language (such as Chinese or Arabic) has many speakers over a wide area and is phonologically diverse.

In the third place, this chapter (as well as comments at the end of the previous chapter) should have made it clear that a phonological analysis of a language is often open to debate. In some instances, linguists may disagree about the number and nature of phonemes in a language. One example already mentioned (in section 3.14 above) is that of the vowel heard in words like *cue* and *few*: if we do indeed take this to be a vowel, the diphthong /ju/, then we must count it as one of the vowel phonemes of English; if we take it to be a sequence of /j/ and /u/, then there is no vowel phoneme /ju/. Another example mentioned earlier in this chapter (section 4.3 above) is the question of whether German has /ç/ and /χ/ as phonemes, rather than /χ/ and /h/ as is usually assumed. There are similar alternative analyses for many languages, which will affect both the total number of phonemes in a language and the nature of the phonemes themselves.

In the fourth place, the very notion of the phoneme makes it difficult to make simple statements. The phoneme is better understood as a point in a system of oppositions, rather than as an item in an inventory. Suppose, for example, that we observe that most, if not all, of the world's languages, including languages as diverse as Aranda, English, Indonesian and Japanese, have a phoneme /t/. What does this actually mean? The apico-alveolar /t/ of Aranda is in contrast with other plosives such as a lamino-dental and a retroflex (or apico-postalveolar); but Aranda has no opposition of voicing, so /t/ is not in contrast with /d/ (and /t/ may sometimes be realized as [d]). On the other hand, English /t/ is opposed to /d/, and its voicelessness is therefore a relevant feature (although in many environments it may be the aspiration of /t/ that is more significant than its voicelessness); but, again unlike Aranda /t/, English /t/ is not in opposition to dental or postalveolar plosives, and, indeed, English /t/ has an apico-dental allophone in a word like *eighth*. Indonesian /t/, like English /t/, is opposed to /d/, but without significant aspiration. Japanese /t/ is different again: it has allophones not found in the other three languages, namely a palatal plosive or affricate before the vowel /i/ and the affricate [ts] before the vowel /u/. In what sense then can we say that these four languages have the same phoneme /t/? It would be more accurate to say that all of these languages exploit – to some extent – a plosive manner of articulation in conjunction with an apico-alveolar place of articulation. But the languages differ in the extent to which they maintain this particular articulatory setting against other options (such as dental articulation) or allow it to be adapted in context (as when English /t/

is realized as dental rather than alveolar, or Japanese /t/ is realized as affricate rather than plosive); and the languages differ also in the way they combine this articulatory setting with other features such as voicing.

With reservations like these in mind, and without trying to compare phonemic inventories that cannot really be compared, it is still possible to make some tentative generalizations.

A pulmonic airstream mechanism (with air coming from the lungs) is normal in the sense that all languages seem to make use of it and some languages use no other airstream. Sounds using other mechanisms – ejectives, implosives, clicks – are relatively uncommon among languages which have been well documented, although they are common in some areas. Ejectives are found among the indigenous languages of the Americas, Africa and Caucasia (in languages such as Armenian and Georgian) and rarely elsewhere. Implosives seem to be relatively common in some parts of Africa but are quite rare elsewhere. Clicks are common in the languages of southern Africa (including Xhosa and Zulu as well as the linguistically distinct Khoisan languages of the Kalahari region) but are virtually unknown outside that area.

All languages seem to make some kind of distinction between consonants and vowels, and most languages have at least a dozen consonant phonemes and at least three vowels. (The smallest phonemic systems in the UPSID database are those of Rotokas, a language spoken on the island of Bougainville, and Mura, one of the indigenous languages of Brazil: Rotokas has six consonants and five vowels, Mura eight consonants and three vowels.)

As mentioned in section 4.2 above, quite a few languages have relatively simple vowel systems. Classical Arabic, Inupik (formerly Eskimo) and many Australian Aboriginal languages have just three vowels, although, as often, it is necessary to qualify this statement. Classical Arabic, for example, has only three vowel qualities, usually represented in Roman transcription as /i/, /a/ and /u/; but vowel length is also distinctive, so that if we include long and short vowels, there are actually six vowels; moreover, the vowel /a/ also combines with a following /w/ or /j/ to create what are in effect the diphthongs /au/ and /ai/; and, in a further elaboration of the basic three-vowel system, modern speakers of Arabic may pronounce these diphthongs as simple vowels (/ai/ as /e/, and /au/ as /o/).

Vowel systems of between five and eight phonemes are common, with five probably the most frequent. Among the better-known languages, Hebrew, Japanese, Modern Greek, Maori, Russian, Spanish and Swahili can all be considered to have a five-vowel system; Indonesian and Romanian have six; Bengali and Italian seven; Javanese and Turkish eight. But again, some of these languages have other distinctive features as well, such as vowel length and nasalization – each of the seven vowels of Bengali, for instance, may be distinctively oral or nasalized – so that it is unwise to dwell on a simple count of vowel phonemes. It is probably fair to say that if a language distinguishes more than about 10 vowels, it is likely to be exploiting diphthongal combinations and additional features, such as length or nasalization, in conjunction with vowel quality. RP English, for example, can be said to have 21 vowel phonemes (as listed in appendix 1.4). But nine of these are clearly diphthongs rather than simple vowels, and five are distinctively long (although not necessarily in contrast with a short vowel of precisely the same quality). Similarly, Thai can be described as having 21 vowel phonemes: but three of these are diphthongs and the other 18 are actually nine pairs of long and short counterparts.

Turning to consonantal articulation, plosives seem to be universal, and fricatives and nasals almost so. Rotokas and Mura, mentioned earlier as UPSID's smallest inventories, both demonstrate that nasal consonants are not universal. The consonants of Rotokas are three voiceless plosives, one voiced plosive, a fricative and a tap; Mura has six plosives (including a glottal stop) and two fricatives. Most languages have fricatives, except in Australia, where the majority of Aboriginal languages do not have any fricative phonemes.

Most of the world's languages seem to have one or more other consonants, using approximant articulation or some other manner such as trill or flap, but no one sound is universal. A trilled or flapped [r], for example, is common but by no means universal: the /r/ phoneme of English (in most of its realizations in most regional varieties) is not a trill or flap; while languages as diverse as Chinese, Inupik and Luganda have an /l/ phoneme but no consonantal r-sound. Many languages have at least one lateral consonant, but Japanese and Tahitian are examples of languages which do not use contrastive lateral articulation, while Korean has [l] and [r] as allophones of a single phoneme. Questions about the occurrence of approximants such as [w] and [j] are particularly difficult to answer because of the scope for alternative analyses (sections 3.11–3.14 above). Taking [w] as an example, we can say that it does occur in English (as in *west* and *woe*) but not in German (orthographic *w* represents [v], not [w], in German); but a language like Spanish has vowel sequences which may or may not be interpreted as containing [w], as in *huevo* 'egg' and *continuo* 'continuous' which may be phonemically represented as /wevo/ and /kontinwo/ (Comrie 1987, pp. 245–6). The status of /w/ is likewise arguable in Italian and Portuguese.

Even the smallest phonemic systems make some use of place of articulation contrasts, and a large number of the world's languages seem to distinguish bilabial and velar from some kind of dental or alveolar place of articulation: thus it is quite common for a language to distinguish /p/ from /t/ from /k/ (where /t/ may be apico-alveolar or apico-dental depending on the language and the phonetic context).

A good number of languages have additional places of articulation for plosives. Probably the most common of these are those usually called retroflex (including apico-postalveolar or sublamino-postalveolar) and palatal. Retroflex plosives are found in many Australian Aboriginal languages and in most of the languages of South Asia (such as Bengali, Pashto, Punjabi, Tamil, Hindi and Urdu). In some languages, including Dutch, Norwegian and some varieties of American English, some speakers may use postalveolar or retroflex plosives after a preceding r-sound. Some Dutch speakers, for example, pronounce the word *hart* 'heart' with a final apico-postalveolar plosive and with little or no articulatory gesture to correspond to the *r* of the spelling: for such speakers, the postalveolar place of articulation may serve as the phonetic realization of alveolar following an (elided) r-sound.

Palatal plosives are very widespread, being virtually universal among Australian Aboriginal languages and occurring elsewhere in languages as diverse as Basque, Hungarian, Indonesian, Thai and Vietnamese. This is again a point for special caution, however, as many languages do not differentiate between palatal plosives and affricates, and there is a close relationship between palatal articulation and affrication. (Notice in English that some occurrences of the affricate /tʃ/ have arisen from assimilation of /t/ to a following lamino-palatal /j/, as in *nature* and *picture*.) Some descriptions of Indonesian or Malay, for example, refer to the initial consonant of words such as *cantik* 'pretty' and *cepat* 'quick' as a voiceless palatal plosive while others identify it as an affricate.

The exploitation of places of articulation is often not uniform across different manners of articulation. English illustrates the point by having, for example, bilabial plosives /p/ and /b/ (but no labio-dental plosives) and labio-dental fricatives /f/ and /v/ (but no bilabial fricatives). Indeed, in the UPSID database there is no record of any language with labio-dental plosives. On the other hand, labio-dental fricatives are very common, much more so than bilabial fricatives. A similar asymmetry is observable with palatal articulation. There are languages, such as Italian and Spanish, which have no palatal plosive phoneme but do have a palatal nasal and a palatal lateral. Nevertheless, some languages are more symmetrical than others (section 4.5 above). It is a striking feature of Australian Aboriginal languages that they tend to have exactly the same places of articulation for plosives and nasals: languages that have five places of articulation for plosives usually have five corresponding nasal consonants, for example.

Voicing is a widespread feature of articulation, although it must be remembered that what appears in a phonemic inventory as /t/ versus /d/ may be realized in various ways: /t/ may be aspirated in some or all environments, /d/ may be only partially voiced, and so on. While a high proportion of languages make some kind of differentiation of this kind, there is a substantial minority of languages in which voicing is not a distinctive feature at all. In the Australian language Warlpiri, for example, there is a single series of plosives (at five points of articulation) which are usually voiceless in word-initial position but may be (partially) voiced in other environments. All other consonants – nasals, laterals, approximants and flaps – are characteristically voiced. Most other Australian Aboriginal languages, at least in the southern half of the country, are similar to Warlpiri in this regard. Other languages which do not exploit voicing can be found among the indigenous languages of the Americas, including Inupik.

There are, of course, languages which distinguish more than two kinds of plosive. Ancient Greek and Thai and some other South-East Asian languages have voiceless aspirated plosives as well as voiceless and voiced. Korean also has a three-way distinction among plosives, but of a somewhat different nature, while many South Asian languages, such as Hindi and Urdu, exploit breathy voicing to create a fourth series of plosives (sometimes referred to as voiced aspirates) alongside voiceless aspirated, voiceless and voiced plosives. (See table 4.5.1 for more details.)

Again, some languages are less consistent or symmetrical than others, and it is not uncommon for a language to have a 'gap' in the way it exploits voicing.

Arabic has no /p/ in contrast with /b/, although it does distinguish /t/ from /d/. Dutch has no /g/, although it does distinguish between /p/ and /b/ and between /t/ and /d/. (The *g* of Dutch words such as *gast* 'guest' and *goed* 'good' represents a fricative, and the voiced plosive [g] occurs only as a conditioned variant of /k/, as in, say, [zagduk] for /zakduk/, *zakdoek* 'handkerchief'.)

Turning to the question of an average number of phonemes, it seems likely that a majority of the world's languages have somewhere between 20 and 40 phonemes. But, as we have said before, the number of phonemes in a language can be altered quite radically by analytical decisions. Suppose, for example, a language has 12 obstruent phonemes, each of which may be distinctively labialized. (Thus we might have /pʷ/ in contrast with /p/, /tʷ/ alongside /t/, and so on, making a total of 24 obstruent phonemes.) But suppose that this language also has the phoneme /w/ and that we decide to analyse the labialized consonants as realizations of obstruent followed by /w/. The number of obstruents is now brought back from 24 to 12. The analysis affects the statistics.

We have already mentioned languages with as few as 11 phonemes. English has 40 or so, the exact number depending on the regional variety being described and on the phonemic analysis itself. Languages can have far more phonemes than this, however, and the largest inventory in the UPSID database has 141 phonemes (Maddieson 1984, p. 7). This is a Khoisan language from southern Africa which has a relatively large number of obstruents and click sounds: among other distinctions, it differentiates voiceless aspirated plosives from voiceless and voiced; it also has distinctively ejective stops; it distinguishes both ejective and aspirated affricates from 'ordinary' voiceless affricates; and it achieves a large array of click sounds by complex articulations such as simultaneous nasalization of clicks and affricated release of clicks.

The examples given above indicate that there are some regional tendencies. Clicks are virtually limited to southern Africa (but they do occur elsewhere, in the secret language of at least one Australian Aboriginal people, and as paralinguistic signs, as in the English use of the click represented as 'tsk tsk' or 'tut tut'). Languages using several implosive consonants seem to be confined to Africa, while languages without any fricative phonemes seem to be found only in Australia. Regional generalizations of this kind are nevertheless rather few in number and of doubtful significance. There are many other cases where similar sounds or patterns of contrast can be found across a range of diverse languages.

In general it is difficult to establish a significant relationship between a language's genetic affiliation and its phonological characteristics. The mere fact that a language's phonological system can change quite substantially over time is enough to show that families of historically related languages do not necessarily share phonological characteristics. To take the example of English, Old English as spoken around a thousand years ago differed phonologically from modern English in a number of ways: it had, for instance, a voiceless velar fricative and front rounded vowels of the kind still heard in German but no longer in modern English; it had no distinction between voiceless and voiced fricatives (the voiceless phonemes having voiced allophones in some environments); and it had distinctive length for both consonants and vowels (with, for instance, a difference in pronunciation between the long [n] of *sunne* 'sun' and

the short [n] of *sunu* 'son'). Thus the phonemic system of Old English looks rather different from that of modern English. To take another example, Ancient Greek had three series of plosives, voiceless aspirated, voiceless and voiced (see table 4.5.1 above), and it had only sibilant fricatives and affricates; changes in pronunciation have been such that Modern Greek now has only voiceless and voiced plosives, but has a much richer series of fricatives than Ancient Greek, including voiceless and voiced labio-dental, dental, palatal and velar fricatives, as well as sibilant /s/ and /z/. Such changes in pronunciation mean that one cannot count on historically related languages to retain phonological similarities.

In summary, generalizations about phonemic inventories should never be taken as bare facts. Hidden behind them lie decisions about which languages have been included and which dialect(s) of the languages have been described, and judgements within the process of making a phonemic analysis and representing the phonological system as a set of phonemes. It is possible to say, tentatively, that some kinds of articulation seem more common than others: vowels and plosives, produced with a pulmonic airstream, seem fundamental, with fricative and nasal consonants also very widespread; many languages also seem to have at least one lateral approximant and some kind of r-sound. Among places of articulation, differentiation of bilabial, dental or alveolar, and velar is very common for plosives, with palatal articulation also widespread. Among fricatives, labio-dental /f/ and a dental or alveolar grooved sibilant /s/ are probably the most common. Far less common – at least among the best-studied languages – are sounds produced other than with air from the lungs, notably ejectives, implosives and clicks, and places of articulation such as uvular and pharyngeal. Voicing (or aspiration) is probably relevant in a majority of languages, but by no means a universally distinctive feature.

The UPSID database, mentioned earlier as a careful sampling of the world's phonemic systems, is explained in Maddieson (1984). Maddieson includes detailed discussion of what inferences can be drawn from the 317 phonemic systems in the database.

Exercises

1 Think of someone you consider to have an easily recognized voice, perhaps a friend or an actor or television personality. Can you mimic the voice? What features of speech production do you think make the voice distinctive?
2 Describe some instances of context-sensitive variation in your own pronunciation of English. Possible examples include the effects of /l/ and /r/ on preceding vowels, or the effects of adjacent consonants on each other, such as /t/ and /r/ in *train* or /d/ and /r/ in *drain*.
3 List the vowel phonemes of your own variety of English. One way to approach this task is to compile a list of minimally contrasting words: for example each of the words *had*, *head*, *heed*, *hid*, *hide*, *hood* . . . has a distinct vowel, as do *den*, *dean*, *din*, *dine*, *done* . . . If you do this with a few different sets of words, adding as many distinct vowels as possible, and then group those which have the same vowel (as *head* and *den*, and *heed* and *dean*, and so on), you should have evidence of the number of distinct vowel phonemes in your speech.

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

   a. allomorph
   b. allophone
   c. archiphoneme
   d. biuniqueness
   e. complementary distribution
   f. elision
   g. free variation
   h. minimal pair
   i. phonemic invariance
   j. phonological conditioning

5 Using the following words of an invented language as data, show that the following pairs of sounds are in complementary distribution: [i] with [e], [u] with [o], [t] with [d] and [s] with [z].

[maza]	[name]	[sana]
[mido]	[nimo]	[size]
[mune]	[nize]	[sude]
[muzo]	[numa]	[tada]
[tane]	[tino]	[tuda]

6 Looking at the following Japanese names in the standard Romanized spelling, what would you conclude about the distribution of the sounds represented by *ch*, *f*, *h*, *s*, *sh*, *t* and *ts*?

   Fuse
   Futatsume
   Hachinohe
   Hashimoto
   Hitachi
   Hofu
   Matsushima
   Misumi
   Shinichi
   Soto
   Susa
   Tate

7 Illustrate the neutralization of phonemic oppositions from English.
8 Explain what is meant by the 'symmetrical patterning' of a phonological system.