

CHAPTER

1 Introduction to speech

CHAPTER OUTLINE

KEY TERMS

*Amplitude**Consonant**International**Phonetic**Alphabet (IPA)**Loudness**Phonetic symbol**Phonetic**transcription**Segment**Syllable**Vowel**Waveform**Writing system*

In this chapter you will learn about: the basic distinction between spoken and written language; the ways in which languages of the world are written; the units from which speech is composed: syllables, vowels and consonants; phonetic symbols as a means of representing speech; speech considered as an acoustic signal; the similarities and differences in the speech sounds used in languages of the world.

Introduction

The way we usually represent and describe speech depends on a powerful idea that is already known by everyone who is literate in a language with an alphabetic writing system. Human listeners can hear speech as a sequence of sounds, and each sound can be represented by a written mark. In this chapter we look at how this idea can be the basis of a comprehensive system of **phonetic symbols**, suitable for representing reliably the sounds of any language – and at how this is different from the many existing writing systems for particular languages.

Sounds and symbols

Although there are estimated to be 5,000 to 8,000 languages in the world, each with its own particular selection of sounds, the total number of symbols required to represent all the sounds of these languages is not very large – it is somewhere around two to three hundred. This, of course, is because many sounds are found again and again in languages. The human speech apparatus, which produces sounds, and the hearing mechanism, which perceives them, are exactly the same all over the world. Languages make their selection from the stock of humanly possible sounds – and some sounds are so common and basic that they are found in almost all languages.











From surveys that compare the sounds employed in hundreds of languages, we know that almost all languages have consonant sounds like those at the beginnings of the English words *tea*, *key*, *pea*, *see*, *fee*, *me* and *knee*, and vowel sounds resembling those heard in *seat* and *sat*. Although English has quite a large system of sounds, none of the sounds is very unusual when seen in a global perspective. The most unusual sounds in English are probably the so-called *th*-sounds heard at the beginning of *think* and *this*. It's interesting that some varieties of English don't use these sounds, but replace them with others that are more frequent in the world's languages. Speakers from London, for instance, often say *fink* rather than *think*.

The International Phonetic Alphabet

The **International Phonetic Alphabet (IPA)** aims to provide a separate symbol for every sound used distinctively in a human language (see Figure 1.1). Using symbols from the IPA, we should be able to represent the pronunciation of any word or phrase in any human language. The IPA has grown and evolved over more than a century of international collaboration, with new symbols being added when new sounds turned up in languages that had not previously been described.

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
	Bilabial		Bilabial	'	Examples:
	Dental		Dental/alveolar	p'	Bilabial
	(Post)alveolar		Palatal	t'	Dental/alveolar
	Palatoalveolar		Velar	k'	Velar
	Alveolar lateral		Uvular	s'	Alveolar fricative

ʌ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
ʋ	Voiced labial-velar approximant	ɭ	Alveolar lateral flap
ɥ	Voiced labial-palatal approximant	ɥ̟	Simultaneous ɥ and X
ħ	Voiceless epiglottal fricative		
ʕ	Voiced epiglottal fricative		Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʡ	Epiglottal plosive		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

o Voiceless	<u>n̥</u> <u>d̥</u>	.. Breathly voiced	<u>b̤</u> <u>a̤</u>	~ Dental	<u>t̪</u> <u>d̪</u>
✓ Voiced	<u>s̹</u> <u>t̹</u>	~ Creaky voiced	<u>b̰</u> <u>a̰</u>	~ Apical	<u>t̺</u> <u>d̺</u>
h Aspirated	<u>tʰ</u> <u>dʰ</u>	~ Linguolabial	<u>t̼</u> <u>d̼</u>	~ Laminal	<u>t̻</u> <u>d̻</u>
More rounded	<u>ɔ̹</u>	W Labialized	<u>tʷ</u> <u>dʷ</u>	~ Nasalized	<u>ẽ</u>
Less rounded	<u>ɔ̜</u>	j Palatalized	<u>tʲ</u> <u>dʲ</u>	n Nasal release	<u>d̪ⁿ</u>
Advanced	<u>u̟</u>	Y Velarized	<u>tʷ</u> <u>dʷ</u>	l Lateral release	<u>d̪ˡ</u>
Retracted	<u>u̠</u>	ʕ Pharyngealized	<u>tˤ</u> <u>dˤ</u>	~ No audible release	<u>d̪̚</u>
Centralized	<u>ẽ</u>	~ Velarized or pharyngealized	<u>ɮ</u>		
× Mid-centralized	<u>ẽ̞</u>	~ Raised	<u>e̞</u> (<u>ɹ̞</u> = voiced alveolar fricative)		
~ Syllabic	<u>n̩</u>	~ Lowered	<u>e̝</u> (<u>β̝</u> = voiced bilabial approximant)		
~ Non-syllabic	<u>e̯</u>	~ Advanced Tongue Root	<u>e̘</u>		
~ Rhoticity	<u>ə̤</u> <u>a̤</u>	~ Retracted Tongue Root	<u>e̙</u>		

Front Central Back

Close i y — ɨ ʉ — ɯ ʊ

Close-mid e ø — ə ɘ — ɤ ʊ

Open-mid ε œ — ɜ ɞ — ʌ ɔ

Open a ɶ — ɑ ɒ

Where symbols appear in pairs, the one to the right represents a rounded vowel.

ˈ	Primary stress	
ˌ	Secondary stress	
ː	Long	ˌfʊnəˈtʃən
ˑ	Half-long	eˑ
ɐ	Extra-short	ɐ
⏏	Minor (foot) group	
⏏⏏	Major (intonation) group	
·	Syllable break	ni.ækt
⏏	Linking (absence of a break)	

ē _{or}	↗	Extra high	ě _{or}	↗	Rising
ē	↗	High	ê	↘	Falling
ē	↔	Mid	ē	↗	High rising
ē	↘	Low	ē	↗	Low rising
ē	↘	Extra low	ē	↗	Rising-falling
↓		Downstep	↗		Global rise
↑		Upstep	↘		Global fall

www.cambridge.org

This is a specimen of Alexander Melville Bell's Visible Speech system, one of many phonetic notation systems invented in the past. Only the IPA enjoyed any lasting success, and it is the only system in use today.

The IPA takes the familiar Latin alphabet as its starting point. The twenty-six letters of the alphabet are all used, mostly with values that seem very natural to us, and further symbols are obtained in a variety of ways:

- by using small capital letters with different meanings from the lower case ones; for example, [g] and [G] stand for different sounds, so do [n] and [N]. Notice how square brackets are put around phonetic symbols.
- by turning or inverting existing letter shapes, as in [ə], [ɹ] [w] [ɤ]
- by using **diacritics**, which are dots, hooks, and other small marks added to symbols, as in [ŋ], [ã]
- by using some letters from the Greek alphabet, such as [φ] and [χ], (though the values attached to the symbols are not necessarily Greek sounds)
- by inventing new shapes, such as [uɰ], [ɳ]. The new shapes mostly look like existing letters, or have some logical feature in them that highlights resemblances among sounds.

It's easy to use IPA symbols in word-processed documents on a computer. For details on how to do it, see the website that accompanies this book.

As you see from Figure 1.1, the chart of the IPA fits on one page. It provides nearly everything that is required to represent speech in any language – at least, in the normal adult pronunciations of those languages. But speakers who have speech disorders sometimes employ types of sound different from those used by any normal speaker. The IPA provides a supplementary chart, and further diacritics, which may be needed to cope with the range of sounds encountered in pathological speech.

After phonetic training, we can listen analytically to words or phrases, imitate them, and record them with symbols. We can start from zero with a language we do not know at all, transcribing words with the help of a co-operative speaker who is willing to repeat words for us and listen to our imitations. For instance, in the first session with a speaker of the language Divehi (Maldivian Islands), the first word asked for was the one for 'hand'. The response was transcribed as [aeʔ]. The two symbols [ae] represent a

particular gliding vowel sound (roughly like that in English *eye*), and [ʔ] represents a **glottal stop**. The native speaker of Divehi would not accept [ae] without the glottal stop as a version of ‘hand’, which showed that she regarded the glottal stop as an essential sound in the word.

The next word asked for was that for ‘finger’. This time the response was noted down as [iŋgiliʔ]. Here the [i] symbols represent vowels approximately like those in English *beat*, and again there is a final glottal stop. The symbol [ŋ] represents a sound which is encountered commonly in English but spelled with *n* in conjunction with other letters (commonly *k* or *g*). It is the nasal sound before the [k] of *think*, or the final sound of *thing* for most speakers. Actually the English word *finger* contains exactly the same sequence [ŋg] noted in the Divehi word. The next word asked for was ‘head’ and this was noted down as [bɔ]. Here the symbol [ɔ] represents a vowel somewhat like that in English *saw*. There seemed to be no glottal stop at the end, but as the other two words noted so far had finished with a glottal stop, the investigator pointed to his head and tried a version [bɔʔ]. This produced laughter from the informant. It turns out that [bɔʔ] is a separate word meaning ‘frog’. This established that the glottal stop was not just an automatic termination for Divehi words, but a significant sound capable of marking distinctions between words.

If this process of imitation and transcription were continued, we could expect to encounter all of the **vowels** and **consonants** employed in the language, and select suitable symbols for transcribing them all. It makes no difference to this process whether the language under investigation happens already to have a written form. As a matter of fact, Divehi is a written language, with a unique alphabet of its own; and the speaker in question was literate in both Divehi and English – but these facts have no relevance to the process of listening, imitating and transcribing.

A glottal stop (or more fully a glottal plosive) is the speech sound that often replaces the [t] of a word such as *better* in a London pronunciation, or the [t] of words such as *football* quite generally in both British and American pronunciation. As we will see, it is formed by a momentary closure of the vocal folds, located in the larynx. Within a word such as *better*, a glottal stop is heard as a brief interval of silence. A glottal stop which follows a vowel, as in this Divehi example, is heard as an abrupt termination of the vowel sound.

Types of transcription

Once a **phonetic transcription** has been made, it should sound right when we read it aloud again. But people often have a wrong idea about how precise phonetic transcription can be. The IPA symbols may look a bit like mathematical symbols, but they are not used with mathematical precision. Learning to transcribe is not at all like learning formal logic or algebra – it is more like learning how to make a recognisable sketch of a face or an object (one kind of transcription is even called impressionistic). Different observers can make somewhat different transcriptions of the same sample of speech, without either of them being necessarily ‘wrong’. And however detailed we make our transcription, it remains only a rough approximation compared with a sound recording, or a film of the speaker’s actions in producing the speech.

Symbols for particular languages

The symbols needed to represent the sounds of any particular language – whether it’s English or Divehi, or any other – will be a selection from

(a subset of) the IPA. As an illustration, symbols for representing one type of English are given in the table below. This is followed by a specimen of transcription for you to try reading.

	Keyword		Keyword
p	pie	j	yes
t	tie	h	hat
k	key	i:	pea
b	buy	ɪ	pit
d	die	e	pet
g	guy	æ	pat
m	my	ɑ:	pa
n	no	ʌ	pup
ŋ	sing	ɒ	pot
f	fee	ɔ:	paw
v	van	ʊ	put
θ	thigh	u:	too
ð	though	ə	soda
s	so	ɜ:	bird
z	zoo	eɪ	pay
ʃ	she	aɪ	pie
ʒ	measure	ɔɪ	boy
tʃ	chip	aʊ	now
dʒ	jam	əʊ	no
w	wet	ɪə	near
r	red	eə	hair
l	let		

The mark ' indicates that the following syllable is stressed, and | indicates a slight pause at the end of a phrase.

ðə 'nɔ:θ 'wind ən ðə 'sʌn wə drɪ'spju:tɪŋ wɪtʃ wəz ðə 'strɒŋgə |
wen ə 'trævlə 'keɪm ə'lon | 'ræpt ɪn ə 'wɔ:m 'kləʊk |
ðeɪ ə'grɪ:d ðət ðə 'wʌn hu: 'fɜ:st sək'sɪ:dɪd ɪn 'meɪkɪŋ ðə 'trævlə
'teɪk ɪz 'kləʊk ɒf | ʃʊd bɪ kən'sɪdəd 'strɒŋgə ðən ðɪ 'ʌðə |
'ðen ðə 'nɔ:θ 'wind 'blu: əz 'hɑ:d əz ɪ: 'kud | bət ðə 'mɔ: hi: 'blu: |
ðə mɔ: 'kləʊslɪ dɪd ðə 'trævlə 'fəʊld ɪz 'kləʊk ə'raʊnd hɪm | ən ət
'lɑ:st ðə 'nɔ:θ 'wind 'geɪv 'ʌp ðɪ ə'tempt | 'ðen ðə 'sʌn 'ʃɒn aʊt
'wɔ:mlɪ | ən ɪ'mɪdʒətɪ ðə 'trævlə 'tʊk 'ɒf ɪz 'kləʊk | ən səʊ ðə 'nɔ:θ
'wind wəz ə'blaɪdʒd tə kən'fes | ðət ðə 'sʌn wəz ðə 'strɒŋgə əv
ðə 'tu: |

(The orthographic version of this text is given at the end of the answers section for this chapter.)

The IPA does not provide fixed transcription systems for particular languages. It provides a stock of symbols, and principles and conventions for using them – but there can be perfectly legitimate differences between transcription systems for one and the same language. As a simple illustration, consider the vowel in a Southern British pronunciation of a word

such as *dress*. It is somewhere between the sounds represented [e] and [ɛ] in the IPA. We could use a diacritic added to one of these symbols to show an intermediate quality – for instance [ɛ̞]. But how inconvenient would it be, all the way through a book on English pronunciation, or in a dictionary that shows pronunciations, to add the diacritic each time? Far better to choose either [e] or [ɛ] for regular use, and give a once-for-all statement that the actual quality is something like [ɛ̞]. In fact, some transcription systems for English use [e] while others use [ɛ].

Syllables

Though both speaker and listener may have the impression that speech is a sequence of sounds, the shortest stretch of speech that a speaker can actually pronounce in a fairly natural way is not the individual sound, but the **syllable**. If a person is asked to speak very slowly, splitting words up into sections (e.g. for dictation), division will usually be into syllables. Thus the word *signal* can be spoken as two chunks separated by a pause: *sig - nal*. This is because *signal* is a word of two syllables. By contrast the word *sign* is a one-syllable word (a **monosyllable**). It cannot be divided into two individually pronounceable parts. It begins like *sigh*, but the remainder is the single sound represented by *n*, which we can only pronounce in an unnatural and disjointed way. The middle part of *sign* is also clearly the same as *I* or *eye*, but removing that portion leaves us with two sounds represented by *s* and *n*, separated by a gap. The conclusion is that *sign*, *sigh* and *eye* are all monosyllables.

A syllable is like one pulse of speech. It always contains one loud or prominent part (almost always a vowel sound), and may optionally have consonant sounds preceding or following the vowel. If we compare the pronunciations of the three syllables *be*, *eat* and *beat* we can hear that they all contain the same vowel sound, which we can represent with its phonetic symbol [i]. In *be*, the vowel is preceded by a consonant sound [b], but nothing comes after the vowel, giving [bi]. In *eat* there is nothing before the vowel, but a consonant sound [t] follows. *Beat* [bit] has both preceding and following consonants.

Segments: vowels and consonants

The term **segment** is another way of referring to the individual speech sounds that make up syllables. Segments are of two kinds: vowels and consonants. Typical vowel segments are [i a u]; a few examples of typical consonants are [m b k f s]. Using V to stand for any vowel and C for any consonant, the structure of a syllable or word can be shown as a string of Vs and Cs. So, for example, the word *book*, pronounced [bʊk], is CVC. This means a sequence of one consonant followed by one vowel which in turn is followed by one consonant. This sort of representation is called a **CV-skeleton**.

Here are some examples of CV-skeletons for English, together with some words that conform to each of them. Remember that we are dealing with pronunciations, not spellings. The double -oo- in the conventional spelling of *book* doesn't mean that the word contains two vowel sounds. To take

V	<i>eye, oh</i>
CV	<i>be, my, see, saw, tea, you</i>
VC	<i>eat, each, aim</i>
CVC	<i>book, chip, thumb, top, win</i>
CCV	<i>draw, glue, stay</i>
CVCCCC	<i>texts</i>
CCCV	<i>straw</i>
CVCVC	<i>unit</i>
CVCCVC	<i>signal</i>
CVVC	<i>going</i>

another example, look at the skeleton for *unit*. This word is spelt with a vowel letter at the beginning, but it is pronounced with a consonant segment in initial position, exactly like the word *you*. Notice also that in *win* and *you* the letters *w* and *y* stand for consonant sounds, whereas in *my*, *saw*, *draw*, *stay*, *straw* they are used as part of the representation of the vowel.

As we see in these examples, when we represent careful pronunciations of whole words with CV-skeletons, there must be one V element for each syllable. The three-syllable word *banana* has the skeleton CVCVCV. In addition to simple vowels, like those heard in *book*, *bit* or *cat*, there are also **diphthongs**, which are vowels of changing or ‘gliding’ quality like those heard in *voice* or *house*. Since *voice* and *house* are one-syllable words, English diphthongs must count as one V element rather than two. Both *voice* and *house* have the structure CVC, rather than CVVC. But *going* has two syllables: one is the stem *go*, which contains a diphthong, the second is the **ending** *-ing*, which contains a further vowel followed by the consonant [ŋ].

Some languages also permit certain consonants to be **syllabic** – that is, to form a syllable by themselves. In English, [l] and [n] may be syllabic, as in the second syllables of certain pronunciations of *settle* [setl̩] or *sudden* [sʌdn̩]. (The IPA diacritic added to the relevant consonants means ‘syllabic’.) These words can be represented CVCC̩. But notice that there are also alternative pronunciations [setəl], [sʌdən] in which the second syllables of these words contain a V element followed by a non-syllabic consonant, giving the structure CVCVC.

Syllables and words

In a sense, spoken words are composed of syllables. The shortest possible words are words of one syllable. The English words *hand*, *arm*, *head*, *eye*, *mouth* are all monosyllables, as are *have*, *be*, *go*, *do*, *make*, *eat*, *die*. Of course, English words may have two, three or more syllables. For example, the English words *mother*, *husband*, *river*, *heaven*, *berry* have two syllables, while *description*, *musical*, *prominent* have three, *applicable* has four, *characteristic* has five, and so on.

But syllables are units of pronunciation rather than elements of word structure. Notice that the word *dog* is a monosyllable, and so is its plural, *dogs*; but the plural clearly consists of two elements of word structure: one is the stem, *dog-*; the other is the ending indicating 'plural'. There are thus two elements of word-structure within the one syllable. On the other hand, *banana* has three syllables, but just one element of word structure (a stem), as the word isn't made up of separate meaningful parts.

Suprasegmentals

Segments aren't the whole story. We also have to pay attention to features that are not themselves segments, and that seem to spread across several successive segments (often a whole syllable). Such properties are called suprasegmentals. **Stress** and **tone** are in this category. In English, for example, the noun *import* and the verb *import* have exactly the same segments. But the words are distinguished according to which of the two syllables is stressed: the noun is IMport but the verb is imPORT. Here we've used capitalisation to give an indication of which syllable is stressed, but as you will see from the chart, the IPA provides symbols for suprasegmentals too, and we will return to these in a later chapter.

Speech as an acoustic signal

Like any sound, speech can be picked up with a microphone, recorded and analysed. Sound is a rapid variation of pressure travelling through some physical medium (such as air). The velocity of sound in air is about 330 metres per second (about 740 miles per hour). When variations in pressure arrive at the eardrum, or at a microphone, they cause vibration (tiny to-and-fro movements) of the eardrum or the diaphragm of the microphone. The human listener experiences hearing a sound, and the microphone produces an electrical signal which can be measured.

A graph showing variation of pressure (or equivalently, movements of the eardrum) as a function of time gives the **waveform** of a sound. In a perfectly quiet place, a microphone picks up no sound, and the resulting waveform will be flat, showing no up and down movement at all. In most locations there will generally be background or **ambient noise** (traffic or aircraft noise from outside, wind, sounds from appliances, and so on) and this will show as constantly fluctuating energy on a waveform. Speaking reasonably close to a microphone results in a waveform that is much bigger than the ambient noise level.

Because the pressure variations in sound are very rapid, the amount of detail we can see in a waveform depends on the scale on which it is plotted. If a lot of time is shown in a small space, the details of the waveform are lost, and we just see blocks of activity.

The **amplitude** of a wave is a measure of the size of the pressure variations (or eardrum movements). The auditory property that is correlated with amplitude is **loudness**. Other things being equal, a wave with larger variation in air pressure will correspond to a louder sound. In Figure 1.2 we can see that the vowels have more energy (are louder) than the consonants.

While English has plenty of basic vocabulary items that are single syllables, not all languages actually permit monosyllabic words. In most indigenous Australian languages, for example, words have to have at least two syllables, following the formula CV(C)CV(C). The brackets show that the consonants in those positions may be present, but are not obligatory. So the formula covers CVCV, CVCCVC, CVCCV and CVCVC.

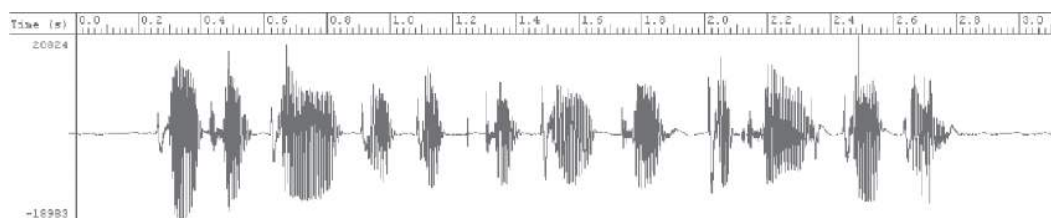


FIGURE 1.2

The waveform for *Peter Piper picked a peck of pickled pepper*. The phrase has been specially chosen so that the twelve syllables of the utterance can be counted. The consonant sounds that separate the syllables have low amplitude, making the prominent sounds at the centre of the syllables easy to see.

Zooming in on the waveform, so that only a small fraction of a second is shown, reveals the detailed pressure variations and can tell us something about the sound of individual segments. It is generally convenient to measure time not in whole seconds, but thousandths of a second. One millisecond (ms) = 1/1000 second.

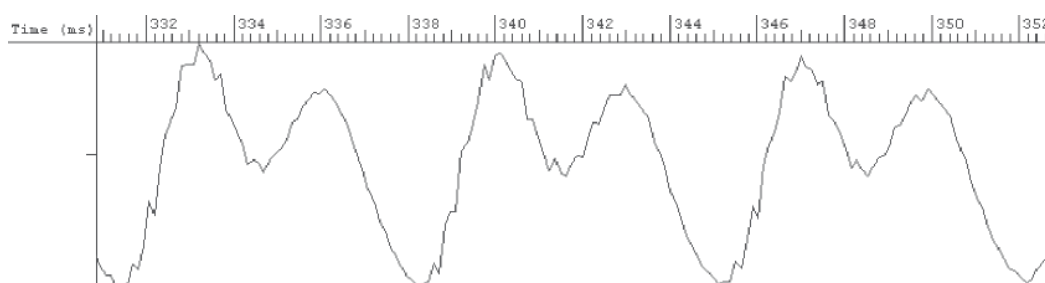


FIGURE 1.3

A short sample (about 22 milliseconds long) from the first vowel of *Peter* in the utterance shown above. The time scale shows milliseconds from the start of the recording. At this point, the waveform has a regular repeating pattern. Listening to this sample, we hear a very short but recognisable vowel [i].

Writing systems

The writing systems used by the languages of the world are many and various and it needs a whole book considerably larger than this one to deal with them in detail. We will give a simple account here, which should give you an idea of the amount of diversity in writing systems. Ways of writing fall into three basic categories: (1) **alphabetic** systems, (2) **syllabaries** and (3) **logographic** systems.

Alphabets

If you are reading this book, you are of course familiar with at least one alphabetic writing system. The writing system of English is a development