

List of Phonetic Symbols and Signs

- a Cardinal Vowel no. 4 (approximately as in French *patte*); used for first element of Eng. diphthong [aɪ]
- æ front vowel between open and open-mid (Eng. vowel in *cat*)
- ɑ Cardinal Vowel no. 5 (approximately as in French *pas*); used for Eng. [ɑ:] in *car*
- ɒ open rounded Cardinal Vowel no. 5 (Eng. vowel in *dog*)
- b voiced bilabial plosive (Eng. *b* in *labour*)
- β voiced ingressive bilabial plosive
- β voiced bilabial fricative
- c voiceless palatal plosive
- ç voiceless palatal fricative
- C Cardinal Vowel
- ɔ Cardinal Vowel no. 6 (approximately as in German *Sonne*); used for Eng. [ɔ:] in *saw*, and first element of diphthong [ɔɪ]
- d voiced alveolar plosive (Eng. *d* in *lady*)
- ɖ voiced ingressive alveolar plosive
- ɗ voiced palato-alveolar affricate
- θ voiced dental fricative (Eng. *th* in *other*)
- e Cardinal Vowel no. 2 (approximately as in French *the*); used for Eng. [e] in *bed*, and first element of diphthong [eɪ]
- ɘ unrounded central vowel (Eng. initial and final vowels in *another*)
- ɚ retroflexed central vowel (American *er* in *water*)
- ɛ Cardinal Vowel no. 3 (approximately as in French *père*); used for first element of diphthong in [eə]
- ɜ unrounded central vowel (Eng. vowel in *bird*)
- ɝ retroflexed central vowel
- f voiceless labiodental fricative (Eng. *f* in *four*)
- ɸ voiced palatal plosive
- ɣ voiced velar plosive (Eng. *g* in *eager*)
- ɶ voiced velar implosive
- ħ voiceless glottal fricative (Eng. *h* in *house*)
- ɦ voiced glottal fricative (sometimes Eng. *h* in *behind*)
- i Cardinal Vowel no. 1 (approximately as in French *sî*); used for Eng. /i:/ in *see*
- ɪ unrounded close central vowel

- i centralized unrounded close-mid vowel (Eng. vowel in *sit*)
- j (unrounded) palatal approximant (Eng. *y* in *you*)
- r voiced alveolar tap (sometimes *r* in Eng. *very*)
- k voiceless velar plosive (Eng. *c* in *car*)
- l voiced alveolar lateral approximant (Eng. *l* in *lay*)
- ɫ voiced alveolar lateral approximant with velarization (Eng. *ll* in *ill*)
- ɬ voiceless alveolar lateral fricative (Welsh *ll*)
- m voiced bilabial nasal (Eng. *m* in *me*)
- ɱ voiced labiodental nasal (Eng. *m* in *comfort*)
- u Cardinal Vowel no. 16 (like Eng. /u:/ with spread lips)
- n voiced alveolar nasal (Eng. *n* in *no*)
- ŋ voiced velar nasal (Eng. *ng* in *sing*)
- ɲ voiced palatal nasal (French *gn* in *vigne*)
- o Cardinal Vowel no. 7 (approximately as in French *eau*)
- ø Cardinal Vowel no. 10 (approximately as in French *peu*)
- œ Cardinal Vowel no. 11 (approximately as in French *peur*)
- θ voiceless dental fricative (Eng. *th* in *thing*)
- p voiceless bilabial plosive (Eng. *p* in *pea*)
- r voiced alveolar trill (an emphatic pronunciation of *r* in Scottish English)
- ɽ voiced post-alveolar approximant (Eng. *r* in *red*)
- ɻ voiced retroflex approximant
- ʀ voiced uvular trill
- ʁ voiced uvular fricative or approximant
- s voiceless alveolar fricative (Eng. *s* in *see*)
- ʃ voiceless palato-alveolar fricative (Eng. *sh* in *she*)
- t voiceless alveolar plosive (Eng. *t* in *tea*)
- tʃ voiceless palato-alveolar affricate
- ɸ dental click
- u Cardinal Vowel no. 8 (approximately as in French *doux*); used for Eng /u:/ in *do*
- ʊ rounded close central vowel
- o centralized rounded close-mid vowel (Eng. *u* in *put*)
- v voiced labiodental fricative (Eng. *v* in *ever*)
- ʌ Cardinal Vowel no. 14; used for Eng. /ʌ/ in *cup*
- ʋ voiced labiodental approximant
- w labial-velar semi-vowel (Eng. *w* in *we*)
- ɱ voiceless labial-velar fricative (sometimes Eng. *wh* in *why*)
- x voiceless velar fricative (Scottish *ch* in *loch*)
- y Cardinal Vowel no. 9 (approximately as in French *du*)
- ʎ voiced palatal lateral approximant (Italian *gl* in *egli*)
- ɣ Cardinal Vowel no. 15
- ɣ voiced velar fricative
- z voiced alveolar fricative (Eng. *z* in *lazy*)
- ʒ voiced palato-alveolar fricative (Eng. *s* in *measure*)
- ɸ voiceless bilabial fricative
- ɸ alveolar lateral click
- ʔ glottal plosive
- : indicates full length of preceding vowel

- indicates half length of preceding vowel
- high unaccented pre-nuclear syllable
- ˆ high falling nuclear tone (and used to indicate primary accent in citation forms)
- ˘ low falling nuclear tone
- ˆ high rising nuclear tone
- ˘ low rising nuclear tone
- ˆ falling-rising nuclear tone
- ˘ rising-falling nuclear tone
- > mid-level nuclear tone
- = stylized tone (high level followed by mid-level)
- ˆ syllable carrying (high) secondary accent
- ˘ syllable carrying (low) secondary accent
- ̃ nasalization, e.g. [õ]
- ˙ centralization, e.g. [ö]
- ˚ more open quality, e.g. [ø]
- ˚ closer quality, e.g. [ɔ]
- ˚ devoiced lenis consonant, e.g. [z̥] (above in the case of [t̥, ʃ̥, ʒ̥])
- ˚ syllabic consonant, e.g. [ŋ̩] (above in the case of [ŋ̩])
- ˚ dental articulation, e.g. [t̪]
- ˚ post-alveolar articulation
- [] phonetic transcription
- // phonemic transcription
- > changed to
- < developed from
- is realized as
- * common in RP (Figs. 8–26 and in Chapter 10)

PART I

Speech and Language

Communication

1.1 *Speech*

One of the chief characteristics of human beings is their ability to communicate to their fellows complicated messages concerning every aspect of their activity. A man possessing the normal human faculties achieves this exchange of information mainly by means of two types of sensory stimulation, auditory and visual. Children learn from a very early age to respond to the sounds and tunes which their elders habitually use in talking to them; and, in due course, from a need to communicate, they begin to imitate the recurrent sound patterns with which they have become familiar. In other words, they begin to make use of speech; and their constant exposure to the spoken form of their own language, together with their need to convey increasingly subtle types of information, leads to a rapid acquisition of the framework of spoken language. Nevertheless, with all the conditions in favour, a number of years pass before they master the sound system used in their community. It is no wonder, therefore, that the learning of another language later in life, acquired artificially in brief and sporadic spells of activity and often without the stimulus arising from an immediate need for communication, will tend to be tedious and rarely more than partially successful. In addition, the more firmly consolidated the basis of a first language becomes and the later in life a second language is begun, the more learners will be subject to resistances and prejudices deriving from the framework of their original language. As we grow older, the acquisition of a new language will normally entail a great deal of conscious, analytical effort, instead of children's ready and facile imitation.

1.2 *Writing*

Later in childhood children will be taught the conventional visual representation of speech—they will learn to use writing. Today, in considering those languages which have long possessed a written form, we are apt to forget that the writing was originally an attempt at reflecting the spoken language, and that the latter precedes the former for both the individual and the community. Indeed, in many languages, so parallel are the two forms felt to be that the written form may be responsible for

changes in pronunciation or may at least tend to impose restraints upon its development. In the case of English, this sense of parallelism, rather than of derivation, may be encouraged by the obvious lack of consistent relationship between sound and spelling. A written form of English, based on the Latin alphabet, has existed for more than 1,000 years and, though the pronunciation of English has been constantly changing during this time, few basic changes of spelling have been made since the fifteenth century. The result is that written English is often an inadequate and misleading representation of the spoken language of today. Clearly it would be unwise, to say the least, to base our judgements concerning the spoken language on prejudices derived from the orthography. Moreover, if we are to examine the essence of the English language, we must make our approach through the spoken rather than the written form. The primary concern of this book will be the production, transmission, and reception of the sounds of English—in other words, the phonetics of English.

1.3 *Language*

From the moment that we abandon orthography as our starting-point, it is clear that the analysis of the spoken form of English is by no means simple. Each of us uses an infinite number of different speech sounds when we speak English. Indeed, it is true to say that it is difficult to produce two sounds which are precisely identical from the point of view of instrumental measurement: two utterances by the same person of the word *cat* may well show quite marked differences when measured instrumentally. Yet we are likely to say that the same sound sequence has been repeated. Additionally we may hear clear and considerable differences of quality in the vowel of *cat* as, for instance, in the London and Manchester pronunciations of the word; yet, though we recognize differences of vowel quality, we are likely to feel that we are dealing with a 'variant' of the 'same' vowel. It seems, then, that we are concerned with two kinds of reality: the concrete, measurable reality of the sounds uttered, and another kind of reality, an abstraction made in our minds, which appears to reduce this infinite number of different sounds to a 'manageable' number of categories. In the first, concrete, approach, we are dealing with sounds in relation to SPEECH; at the second, abstract, level, our concern is the behaviour of sounds in a particular LANGUAGE. A language is a system of conventional signals used for communication by a whole community. This pattern of conventions covers a system of significant sound units (the PHONEMES), the inflexion and arrangement of 'words', and the association of meaning with words. An utterance, an act of speech, is a single concrete manifestation of the system at work. As we have seen, several utterances which are plainly different on the concrete, phonetic, level may fulfil the same function, i.e. are the 'same', on the systematic language level. It is important in any analysis of spoken language to keep this distinction in mind and we shall later be considering in some detail how this dual approach to the utterance is to be made. It is not, however, always possible or desirable to keep the two levels of analysis entirely separate: thus, as we shall see, we will draw upon our knowledge of the linguistically significant units to help us in determining how the speech continuum shall be divided up on the concrete, phonetic, level; and again, our classification of linguistic units will be helped by our knowledge of their phonetic features.

1.4 *Redundancy*

Finally, it is well to remember that, although the sound system of our spoken languages serves us primarily as a medium of communication, its efficiency as such an instrument of communication does not depend upon the perfect production and reception of every single element of speech. A speaker will, in almost any utterance, provide the listener with far more cues than he needs for easy comprehension. In the first place, the situation, or context, will itself delimit very largely the purport of an utterance. Thus, in any discussion about a zoo, involving a statement such as 'We saw the lions and tigers', we are predisposed by the context to understand *lions*, even if the *n* is omitted and the word actually said is *liars*. Or again, we are conditioned by grammatical probabilities, so that a particular sound may lose much of its significance; e.g. in the phrase 'These men are working', the quality of the vowel in *men* is not as vitally important for deciding whether it is a question of *men* or *man* as it would be if the word were said in isolation, since here the plurality is determined in addition by the demonstrative adjective preceding *men* and the verb form following. Then again, there are particular probabilities in every language as to the different combinations of sounds which will occur. Thus in English, if we hear an initial *th* sound [ð], we expect a vowel to follow, and of the vowels some are much more likely than others. We distinguish such sequences as *-gt* and *-dl* in final positions, e.g. in *beagle* and *beadle*, but this distinction is not relevant initially, so that even if *dloves* is said, we understand *gloves*. Or again, the total rhythmic shape of a word may provide an important cue to its recognition: thus, in a word such as *become*, the general rhythmic pattern may be said to contribute as much to the recognition of the word as the precise quality of the vowel in the first, weakly accented, syllable. Indeed, we may come to doubt the relative importance of vowels as a help to intelligibility, since we can replace our twenty English vowels by the single vowel [ə] in any utterance and still, if the rhythmic pattern is kept, retain a high degree of intelligibility. An utterance, therefore, will provide a large complex of cues for the listener to interpret, but a great deal of this information will be REDUNDANT, as far as the listener's needs are concerned. On the other hand, such an over-proliferation of cues will serve to offset any disturbance such as noise or to counteract the sound-quality divergences which may exist between speakers of two dialects of the same language. But to insist, for instance, upon exaggerated articulation in order to achieve clarity may well be to go beyond the requirements of speech as a means of communication; indeed, certain obscurations of quality are, and have been for many centuries, characteristic of English. Aesthetic judgements on speech, such as those which deplore the use of the 'intrusive *r*', take into account social considerations of a somewhat different order from those involved in a study of speech as communication.

1.5 *Phonetics and Linguistics*

This book describes the sound system of English, but it should be remembered that such a description forms only part of the total description of a language. A complete description of the current state of a language provides information on a number of interrelated components.

The PHONETICS of a language concerns the concrete characteristics (articulatory, acoustic, auditory) of the sounds used in languages, while PHONOLOGY concerns how sounds function in a systemic way in a particular language. The traditional approach to phonology is through PHONEMICS, which analyses the stream of speech into a sequence of contrastive segments, 'contrastive' here meaning 'contrasting with other segments which might change the meaning' (see further §5.3 below). The phonemic approach to phonology is not the only type of phonological theory but it is the most accessible to those with no training in linguistic theory, besides being more relatable to the writing system. Hence the major part of this book is set within phonemic analysis. Besides being concerned with the sounds of a language, both phonetics and phonology must also describe the combinatory possibilities of the sounds (the PHONOTACTICS OF SYLLABLE STRUCTURE) and the PROSODY of the language, that is, how features of pitch, loudness, and length work to produce accent, rhythm, and intonation. Additionally, a study can be made of the relationship between the sounds of a language and the letters used in its writing system (GRAPHOLOGY OR GRAPHEMICS).

While this book presents a detailed description of the phonetics and phonemics of English, reference will need to be made from time to time to other components of the language:

(1) The LEXICON—the words of the language, the sequence of phonemes of which they are composed, together with their meanings.

(2) The MORPHOLOGY—the structure of words, in particular their inflexion (e.g. *start/started*—here the past-tense morpheme is added to the stem morpheme). Statements can be made of the phonemic structure of morphemes—the MORPHOPHONEMICS. So the morphophonemics of the English plural morpheme involve the MORPHOPHONEMIC ALTERNATIONS illustrated by the /s/ in *cats*, the /z/ in *dogs*, and the /ɪz/ in *losses*.

(3) The SYNTAX—the description of categories like noun and verb, and the system of rules governing the structure of phrases, clauses, and sentences in terms of order and constituency.

(4) The SEMANTICS—the meaning of words and the relationship between word meanings, and the way such meanings are combined to give the meanings of sentences.

(5) The PRAGMATICS—the influence of situation on the interpretation of utterances.

Moreover various other aspects of linguistics will involve phonetics and phonology. STYLISTICS concerns the variations involved in different situations and in different styles of speech. SOCIOLINGUISTICS concerns the interaction between language and society (e.g. the variation involved across classes and between the sexes). DIALECTOLOGY (often considered a branch of sociolinguistics) concerns the variation in the same language in different regions. PSYCHOLINGUISTICS concerns the behaviour of human beings in their production and perception of language (e.g. how far do we plan ahead and how much of an utterance do we decode at a time?). LANGUAGE ACQUISITION concerns children's learning of their first language, whereas APPLIED LINGUISTICS principally concerns the acquisition of a second language.

Finally, it is clear that the various components of a language are always

undergoing change in time. The state of a language at any (SYNCHRONIC) moment must be seen against a background of its historical (DIACHRONIC) evolution. It is for this reason that this book includes information on earlier states of the sound system of English, with some speculation on possible developments in the future.

2

The Production of Speech: The Physiological Aspect

2.1 The Speech Chain

Any manifestation of language by means of speech is the result of a highly complicated series of events. The communication in sound of such a simple concept as 'It's raining' involves a number of activities on the part of the speaker. In the first place, the formulation of the concept will take place at a linguistic level, i.e. in the brain; the first stage may, therefore, be said to be psychological. The nervous system transmits this message to the so-called 'organs of speech' and these in turn behave in a conventional manner, which, as we have learned by experience, will have the effect of producing a particular pattern of sound; the second important stage for our purposes may thus be said to be articulatory or physiological. The movement of our organs of speech will create disturbances in the air, or whatever the medium may be, through which we are talking; these varying air pressures may be investigated and they constitute the third stage in our chain, the physical, or acoustic. Since communication generally requires a listener as well as a speaker, these stages will be reversed at the listening end: the reception of the sound waves by the hearing apparatus (physiological) and the transmission of the information along the nervous system to the brain, where the linguistic interpretation of the message takes place (psychological). Phonetic analysis has often ignored the role of the listener. But any investigation of speech as communication must ultimately be concerned with both the production and the reception ends.

Our immediate concern, however, is with the speaker's behaviour and more especially, on the concrete speech level, with the activity involved in the production of sounds. For this reason, we must now examine the articulatory stage (the speech mechanism) to discover how the various organs behave in order to produce the sounds of speech.

2.2 The Speech Mechanism

Man possesses, in common with many other animals, the ability to produce sounds by using certain of his body's mechanisms. The human being differs from other animals in that he has been able to organize the range of sounds which he can emit into a highly efficient system of communication. Non-human animals rarely progress beyond the stage of using the sounds they produce as a reflex of certain basic stimuli to signal fear, hunger, sexual excitement, and the like. Nevertheless, like other animals, man when he speaks makes use of organs whose primary physiological function is unconnected with vocal communication; in particular, those situated in the respiratory tract.

2.2.1 Sources of Energy: The Lungs

The most usual source of energy for our vocal activity is provided by an airstream expelled from the lungs. There are languages which possess sounds not requiring lung (pulmonic) air for their articulation, and, indeed, in English we have one or two extralinguistic sounds, such as the one we write as *tut-tut* and the noise of encouragement made to horses, which are produced without the aid of the lungs; but all the essential sounds of English use lung air for their production. Our utterances are, therefore, largely shaped by the physiological limitations imposed by the capacity of our lungs and by the muscles which control their action. We are obliged to pause in articulation in order to refill our lungs with air, and the number of energetic peaks of exhalation which we make will to some extent condition the division of speech into sense-groups. In those cases where the airstream is not available for the upper organs of speech, as when, after the removal of the larynx, lung air does not reach the mouth but escapes from an artificial aperture in the neck, a new source of energy, such as stomach air, has to be employed; a new source of this kind imposes restrictions of quite a different nature from those exerted by the lungs, so that the organization of the utterance into groups is changed and variation of energy is less efficiently controlled.

A number of techniques are available for the investigation of the activity in speech of the lungs and their controlling muscles. At one time air pressure within the lungs was observed by the reaction of an air-filled balloon in the stomach. On the basis of such evidence from a gastric balloon, it was at one time claimed that syllables were formed by chest pulses.¹ Such a primitive procedure was replaced by the technique of electromyography, which demonstrated the electrical activity of those respiratory muscles most concerned in speech, notably the internal intercostals; this technique disproved the relationship between chest pulses and syllables.² X-ray photography can reveal the gross movements of the ribs and hence by inference the surrounding muscles, although the technique of Magnetic Resonance Imaging (MRI) is now preferred on medical grounds.

¹ Stetson (1951).

² Ladefoged (1967).

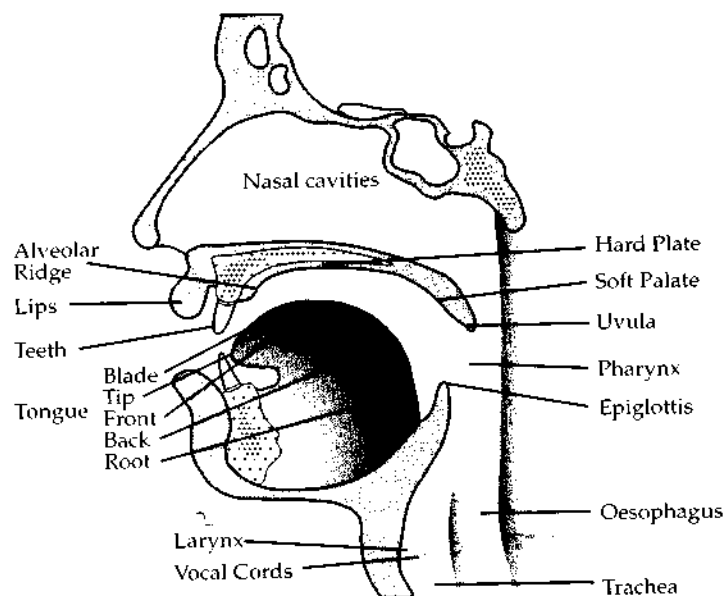


Fig. 1. Organs of speech.

2.2.2 The Larynx and Vocal Folds

The airstream provided by the lungs undergoes important modifications in the upper parts of the respiratory tract before it acquires the quality of a speech sound. First of all, in the TRACHEA or windpipe, it passes through the LARYNX, containing the so-called VOCAL FOLDS, often, less correctly, called the vocal cords, or even vocal chords (see Fig. 1).

The larynx is a casing, formed of cartilage and muscle, situated in the upper part of the trachea. Its forward portion is prominent in the neck below the chin and is commonly called the 'Adam's apple'. Housed within this structure from back to front are the vocal folds, two folds of ligament and elastic tissue which may be brought together or parted by the rotation of the arytenoid cartilages (attached at the posterior end of the folds) through muscular action. The inner edge of these folds is typically about 17 to 22 mm long in males and about 11 to 16 mm in females.³ The opening between the folds is known as the GLOTTIS. Biologically, the vocal folds act as a valve which is able to prevent the entry into the trachea and lungs of any foreign body, or which may have the effect of enclosing the air within the lungs to assist in muscular effort on the part of the arms or the abdomen. In using the vocal folds for speech, the human being has adapted and elaborated upon this original open-or-shut function in the following ways (see Fig. 2).

(1) The glottis may be held tightly closed, with the lung air pent up below it. This 'glottal stop' [ʔ] frequently occurs in English, e.g. when it precedes the energetic articulation of a vowel as in *apple* [ʔæp] or when it reinforces /p,t,k/ as in *clock*

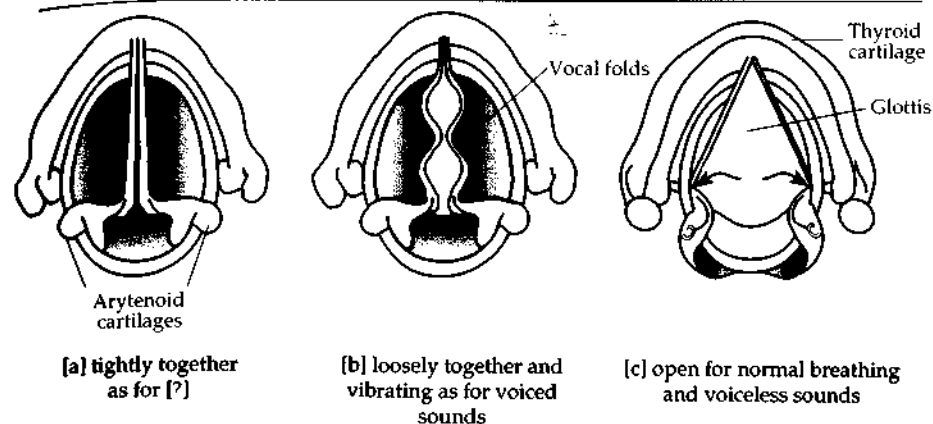


Fig. 2. The vocal cords as seen from above.

[kɒʔk] or even replaces them, as in *cotton* [kɒʔŋ]. It may also be heard in defective speech, such as that arising from cleft palate, when [ʔ] may be substituted for the stop consonants, which, because of the nasal air escape, cannot be articulated with proper compression in the mouth cavity.

(2) The glottis may be held open as for normal breathing and for voiceless sounds like [s] in *sip* and [p] in *peak*.

(3) The action of the vocal folds which is most characteristically a function of speech consists in their role as a vibrator set in motion by lung air—the production of voice, or phonation; this vocal-fold vibration is a normal feature of all vowels or of such a consonant as [z] compared with voiceless [s]. In order to achieve the effect of voice, the vocal folds are brought sufficiently close together that they vibrate when subjected to air pressure from the lungs. This vibration, of a somewhat undulatory character, is caused by compressed air forcing the opening of the glottis and the resultant reduced air pressure permitting the elastic folds to come together once more; the vibratory effect may easily be felt by touching the neck in the region of the larynx or by putting a finger over each ear flap when pronouncing a vowel or [z] for instance. In the typical speaking voice of a man, this opening and closing action is likely to be repeated between 100 and 150 times in a second, i.e. there are that number of cycles of vibration (called Hertz, which is abbreviated to Hz); in the case of a woman's voice, this frequency of vibration might well be between 200 and 325 Hz. We are able, within limits, to vary the speed of vibration of our vocal folds or, in other words, are able consciously to change the pitch of the voice produced in the larynx; the more rapid the rate of vibration, the higher is the pitch (an extremely low rate of vibration being partly responsible for what is usually called creaky voice). Normally the vocal folds come together rapidly and part more slowly, the opening phase of each cycle thus being longer than the closing phase. This gives rise to 'modal' (or 'normal') voice which is used for most of English speech. Other modes of vibration result in other voice qualities, most notably breathy and creaky voice, which are used contrastively in a number of languages. (See also §5.8.) Moreover, we are able, by means of variations in pressure from the lungs, to modify the size of the puff of air which escapes at each vibration of the vocal folds; in other words, we can alter the amplitude of the vibration, with a corresponding change of

³ Clark and Yallop (1990)

loudness of the sound heard by a listener. The normal human being soon learns to manipulate his glottal mechanism so that most delicate changes of pitch and loudness are achieved. Control of this mechanism is, however, very largely exercised by the ear, so that such variations are exceedingly difficult to teach to those who are born deaf, and a derangement of pitch and loudness control is liable to occur among those who become totally deaf later in life.

(4) One other action of the larynx should be mentioned. A very quiet whisper may result merely from holding the glottis in the voiceless position. But the more normal whisper, by means of which we are able to communicate with some ease, can be felt to involve energetic articulation and considerable stricture in the glottal region. Such a whisper may in fact be uttered with an almost total closure of the glottis and an escape of air in the region of the arytenoids.

The simplest way of observing the behaviour of the vocal cords is by the use of a laryngoscope, which gives a stationary mirrored image of the glottis. Using stroboscopic techniques, it is possible to obtain a moving record, and high-speed films have been made of the vocal cords, showing their action in ordinary breathing, producing voice and whisper, and closed as for a glottal stop. The modern technique of observation is to use fiberoptic endoscopy coupled if required with a videocamera.

2.2.3 The Resonating Cavities

The airstream, having passed through the larynx, is now subject to further modification according to the shape assumed by the upper cavities of the pharynx and mouth, and according to whether the nasal cavity is brought into use or not. These cavities function as the principal resonators of the voice produced in the larynx.

2.2.3.1 The Pharynx The pharyngeal cavity (see Fig.1) extends from the top of the trachea and oesophagus, past the epiglottis and the root of the tongue, to the region at the rear of the soft palate. It is convenient to identify these sections of the pharynx by naming them: laryngopharynx, oropharynx, nasopharynx. The shape and volume of this long chamber may be considerably modified by the constrictive action of the muscles enclosing the pharynx, by the movement of the back of the tongue, by the position of the soft palate which may, when raised, exclude the nasopharynx, and by the raising of the larynx itself. The position of the tongue in the mouth, whether it is advanced or retracted, will affect the size of the oropharyngeal cavity; the modifications in shape of this cavity should, therefore, be included in the description of any vowel. It is a characteristic of some kinds of English pronunciation that certain vowels, e.g. the [æ] vowel in *sad*, are articulated with a strong pharyngeal contraction; in addition, a constriction may be made between the lower rear part of the tongue and the wall of the pharynx so that friction, with or without voice, is produced, such fricative sounds being a feature of a number of languages.

The pharynx may be observed by means of a laryngoscope or fiberoptic nasendoscopy, and its constrictive actions are revealed by lateral x-ray photography or, nowadays, preferably by MRI.

The escape of air from the pharynx may be effected in one of three ways:

(1) The soft palate may be lowered, as in normal breathing, in which case the air may escape through the nose and the mouth. This is the position taken up by the soft palate in articulation of the French nasalized vowels in such a phrase as *un bon vin blanc* [œ bõ vë blã], the particular quality of such vowels being achieved through the function of the nasopharyngeal cavity. Indeed, there is no absolute necessity for nasal airflow out of the nose, the most important factor in the production of nasality being the sizes of the posterior oral and nasal openings (some speakers may even make the nasal cavities vibrate through nasopharyngeal mucus or through the soft palate itself).⁴

(2) The soft palate may be lowered so that a nasal outlet is afforded to the airstream, but a complete obstruction is made at some point in the mouth, with the result that, although air enters all or part of the mouth cavity, no oral escape is possible. A purely nasal escape of this sort occurs in such nasal consonants as [m,n,ŋ] in the English words *ram*, *ran*, *rang*. In a snore and some kinds of defective speech, this nasal escape may be accompanied by friction between the rear side of the soft palate and the pharyngeal wall.

(3) The soft palate may be held in its raised position, eliminating the action of the nasopharynx, so that the air escape is solely through the mouth. All normal English sounds, with the exception of the nasal consonants mentioned, have this oral escape. Moreover, if for any reason the lowering of the soft palate cannot be effected, or if there is an enlargement of the organs enclosing the nasopharynx or a blockage brought about by mucus, it is often difficult to articulate either nasalized vowels or nasal consonants. In such speech, typical of adenoidal enlargement or the obstruction caused by a cold, the French phrase mentioned above would have its nasalized vowels turned into their oral equivalents and the English word *morning* would have its nasal consonants replaced by [b,d,g]. On the other hand, an inability to make an effective closure by means of the raising of the soft palate—either because the soft palate itself is defective or because an abnormal opening in the roof of the mouth gives access to the nasal cavity—will result in the general nasalization of vowels and the failure to articulate such oral stop consonants as [b,d,g]. This excessive nasalization (or hypernasality) is typical of such a condition as cleft palate.

It is evident that the action of the soft palate is accessible to observation by direct means, as well as by lateral x-ray photography and MRI; the pressure of the air passing through the nasal cavities may be measured at the nostrils or within the cavities themselves.

2.2.3.2 The Mouth Although all the cavities so far mentioned play an essential part in the production of speech sounds, most attention has traditionally been paid to the behaviour of the cavity formed by the mouth. Indeed, in many languages the word *tongue* is used to refer to our speech and language activity. Such a preoccupation with the oral cavity is doubtless due to the fact that it is the most readily accessible and easily observed section of the vocal tract; but there is in such an attitude a danger of gross oversimplification. Nevertheless, it is true that the shape of the mouth determines finally the quality of the majority of our speech sounds.

⁴ Laver (1980).

Far more finely controlled variations of shape are possible in the mouth than in any other part of the speech mechanism.

The only boundaries of this oral chamber which may be regarded as relatively fixed are, in the front, the teeth; in the upper part, the hard palate; and, in the rear, the pharyngeal wall. The remaining organs are movable: the lips, the various parts of the tongue, and the soft palate with its pendant uvula (see Fig. 1). The lower jaw, too, is capable of very considerable movement; its movement will control the gap between the upper and lower teeth and also to a large extent the disposition of the lips. The space between the upper and lower teeth will often enter into our description of the articulation of sounds; in all such cases, it is clear that the movement of the lower jaw is ultimately responsible for the variation described. Movement of the lower jaw is also one way of altering the distance between the tongue and the roof of the mouth.

It is convenient for our descriptive purposes to divide the roof of the mouth into three parts: moving backwards from the upper teeth, first, the teeth ridge (adjective: ALVEOLAR), which can be clearly felt behind the teeth; secondly, the bony arch which forms the hard palate (adjective: PALATAL), which varies in size and arching from one individual to another; and finally, the soft palate (adjective: VELAR), which, as we have seen, is capable of being raised or lowered, and at the extremity of which is the uvula (adjective: UVULAR). All these parts can be readily observed by means of a mirror.

(1) Of the movable parts, the lips (adjective: LABIAL) constitute the final orifice of the mouth cavity whenever the nasal passage is shut off. The shape which they assume will, therefore, affect very considerably the shape of the total cavity. They may be shut or held apart in various ways. When they are held tightly shut, they form a complete obstruction or occlusion to the airstream, which may either be momentarily prevented from escaping at all, as in the initial sounds of *pat* and *bat*, or may be directed through the nose by the lowering of the soft palate, as in the initial sound of *mat*. If the lips are held apart, the positions they assume may be summarized under five headings:

(a) held sufficiently close together over all their length that friction occurs between them. Fricative sounds of this sort, with or without voice, occur in many languages and the voiced variety [β] is sometimes wrongly used by foreign speakers of English for the first sound in the words *vet* or *wet*;

(b) held sufficiently far apart for no friction to be heard, yet remaining fairly close together and energetically spread. This shape is taken up for vowels like that in *see* and is known as the SPREAD lip position;

(c) held in a relaxed position with a lowering of the lower jaw. This is the position taken up for the vowel of *get* and is known as the NEUTRAL position;

(d) tightly pursed, so that the aperture is small and rounded, as in the vowel of *do*, or more markedly so in the French vowel of *doux*. This is the CLOSE ROUNDED position;

(e) held wide apart, but with slight projection and rounding, as in the vowel of *got*. This is the OPEN ROUNDED position.

Variations of these five positions may be encountered, e.g. in the vowel of *saw*, for which a type of lip-rounding between open and close is commonly used. It will be seen from the examples given that lip position is particularly significant in the

formation of vowel quality. English consonants, on the other hand, with the exception of [p,b,m,w], whose primary articulation involves lip action, will tend to share the lip position of the adjacent vowel. In addition, the lower lip is an active articulator in the pronunciation of [f,v], a light contact being made between the lower lip and the upper teeth.

(2) Of all the movable organs within the mouth, the tongue is by far the most flexible, and is capable of assuming a great variety of positions in the articulation of both vowels and consonants. The tongue is a complex muscular structure which does not show obvious sections; yet, since its position must often be described in considerable detail, certain arbitrary divisions are made. When the tongue is at rest, with its tip lying behind the lower teeth, that part which lies opposite the hard palate is called the FRONT and that which faces the soft palate is called the BACK, with the region where the front and back meet known as the CENTRE (adjective: CENTRAL). These areas together with the root are sometimes collectively referred to as the body of the tongue. The tapering section facing the teeth ridge is called the blade (adjective: LAMINAL) and its extremity the tip (adjective: APICAL). The edges of the tongue are known as the rims.

Generally, in the articulation of vowels, the tongue tip remains low behind the lower teeth. The body of the tongue may, however, be 'bunched up' in different ways, e.g. the front may be the highest part, as when we say the vowel of *he*; or the back may be most prominent, as in the case of the vowel in *who*; or the whole surface may be relatively low and flat, as in the case of the vowel in *ah*. Such changes of shape can be felt if the above words are said in succession. These changes, moreover, together with the variations in lip position, have the effect of modifying very considerably the size of the mouth cavity and of dividing this chamber into two parts: that cavity which is in the forward part of the mouth behind the lips and that which is in the rear, in the region of the pharynx.

The various parts of the tongue may also come into contact with the roof of the mouth. Thus, the tip, blade, and rims may articulate with the teeth, as for the *th* sounds in English, or with the upper alveolar ridge, as in the case of /t,d,s,z,n/, or the apical contact may be only partial, as in the case of /l/ (where the tip makes firm contact whilst the rims make none), or intermittent in a trilled /r/ as in some forms of Scottish English. In some languages, notably those of India, Pakistan, and Sri Lanka, the tip contact may be retracted to the very back of the teeth ridge or even slightly behind it; the same kind of retroflexion, without the tip contact, is typical of some kinds of English /r/, e.g. those used in south-west England and in the USA.

The front of the tongue may articulate against or near to the hard palate. Such a raising of the front of the tongue towards the palate (palatalization) is an essential part of the [ʃ,ʒ] sounds in English words such as *she* and *measure*, being additional to an articulation made between the blade and the alveolar ridge; or again, it is the main feature of the [j] sound initially in *yield*.

The back of the tongue can form a total obstruction by its contact with the soft palate, raised in the case of [k,g] and lowered for [ŋ], as in *sing*; or again, there may merely be a narrowing between the soft palate and the back of the tongue, so that friction of the type occurring finally in the Scottish pronunciation of *loch* is heard. And finally, the uvula may vibrate against the back of the tongue, or there may be a

narrowing in this region which causes uvular friction, as at the beginning of the French word *rouge*.

It will be seen from these few examples that, whereas for vowels the tongue is generally held in a position which is convex in relation to the roof of the mouth, some consonant articulations, such as the southern British English /r/ in *red* and the /l/ in *table*, will involve the 'hollowing' of the body of the tongue so that it has, at least partially, a concave relationship with the roof of the mouth.

Moreover, the surface of the tongue, viewed from the front, may take on various forms: there may be a narrow groove running from back to front down the mid line as for the /s/ in *see*, or the grooving may be very much more diffuse as in the case of the /ʃ/ in *ship*; or again, the whole tongue may be laterally contracted, with or without a depression in the centre (sulcalization), as is the case with various kinds of *r* sounds.

(3) The oral speech mechanism is readily accessible to direct observation as far as the lip movements are concerned, as are many of the tongue movements which take place in the forward part of the mouth. A lateral view of the shape of the tongue over all its length and its relationship with the palate and the velum may be obtained by means of still and moving x-ray photography and by MRI. It is not, however, to be expected that pictures of the articulation of, say, the vowel in *cat* will show an identical tongue position for the pronunciation of a number of individuals. Not only is the sound itself likely to be different from one individual to another, but, even if the sound is for all practical purposes the 'same', the tongue positions may be different, since the boundaries of the mouth cavity are not identical for two speakers; and, in any case, two sounds judged to be the same may be produced by the same individual with different articulations. When, therefore, we describe an articulation in detail, it should be understood that such an articulation is typical for the sound in question, but that variations are to be expected.

Palatography, showing the extent of the area of contact between the tongue and the roof of the mouth, has long been a more practical and informative way of recording tongue movements. At one time the palate was coated with a powdery substance, the articulation was made, and the 'wipe-off' subsequently photographed. But the modern method uses electropalatography, whereby electrodes on a false palate respond to any tongue contact, the contact points being simultaneously registered on a visual display. This has the advantage of showing a series of representations of the changing contacts between the tongue and the palate during speech. Electropalatograms of this sort are used to illustrate the articulations of consonants in Chapter 9.

2.3 *Articulatory Description*

We have now reviewed briefly the complex modifications which are made to the original airstream by a mechanism which extends from the lungs to the mouth and nose. The description of any sound necessitates the provision of certain basic information:

- (1) The nature of the airstream; usually, this will be expelled by direct action of the lungs, but we shall later consider cases where this is not so.
- (2) The action of the vocal folds; in particular, whether they are closed, wide apart, or vibrating.

(3) The position of the soft palate, which will decide whether or not the sound has nasal resonances.

(4) The disposition of the various movable organs of the mouth, i.e. the shape of the lips and tongue, in order to determine the nature of the related oral and oropharyngeal cavities.

In addition, it may be necessary to provide other information concerning, for instance, a particular secondary narrowing, or tenseness which may accompany the primary articulation; or again, when it is a question of a sound with no steady state to describe, an indication of the kind of movement which is taking place. A systematic classification of possible speech sounds is given in Chapter 4.

3

The Sounds of Speech: The Acoustic and Auditory Aspects

3.1 Sound Quality

To complete an act of communication, it is not normally sufficient that our speech mechanism should simply function in such a way as to produce sounds; these in turn must be received by a hearing mechanism and interpreted, after having been transmitted through a medium, such as the air, which is capable of conveying sounds. We must now, therefore, examine briefly the nature of the sounds which we hear, the characteristics of the transmission phase of these sounds, and the way in which these sounds are perceived by a listener.

When we listen to a continuous utterance, we perceive an ever-changing pattern of sound. As we have seen, when it is a question of our own language, we are not conscious of all the complexities of pattern which reach our ears: we tend consciously to perceive and interpret only those sound features which are relevant to the intelligibility of our language. Nevertheless, despite this linguistic selection which we ultimately make, we are aware that this changing pattern consists of variations of different kinds: of SOUND QUALITY—we hear a variety of vowels and consonants; of PITCH—we appreciate the melody, or intonation, of the utterance; of LOUDNESS—we will agree that some sounds or syllables sound 'louder' than others; and of LENGTH—some sounds will be appreciably longer to our ears than others. These are judgements made by a listener in respect of a sound continuum emitted by a speaker and, if the sound stimulus from the speaker and response from the listener are made in terms of the same linguistic system, then the utterance will be meaningful for speaker and listener alike. It is reasonable to assume, therefore, that there is some constant relationship between the speaker's articulation and the listener's reception of sound variations. In other words, it should be possible to link through the transmission phase the listener's impressions of changes of quality, pitch, loudness, and length to some articulatory activity on the part of the speaker. It will in fact be seen that an exact parallelism or correlation between the production, transmission, and reception phases of speech is not always easy to establish, the investigation of such relationships being one of the tasks of present-day phonetic studies.

The formation of any sound requires that a vibrating medium should be set in motion by some kind of energy. We have seen that in the case of the human speech mechanism the function of vibrator is often fulfilled by the vocal folds, and that these are activated by air pressure from the lungs. In addition, any such sound produced in the larynx is modified by the resonating chambers of the pharynx, mouth, and, in certain cases, the nasal cavities. The listener's impression of sound quality will be determined by the way in which the speaker's vibrator and resonators function together.

Speech sounds, like other sounds, are conveyed to our ears by means of waves of compression and rarefaction of the air particles (the commonest medium of communication). These variations in pressure, initiated by the action of the vibrator, are propagated in all directions from the source, the air particles themselves vibrating at the same rate (or frequency) as the original vibrator. In speech, these vibrations may be of a complex but regular pattern, producing 'tone' such as may be heard in a vowel sound; or they may be of an irregular kind, producing 'noise', such as we have in the consonant /s/; or there may be both regular and irregular vibrations present, i.e. a combination of tone and noise, as in /z/. In the production of normal vowels, the vibrator is normally provided by the vocal folds; in the case of many consonant articulations, however, a source of air disturbance is provided by constriction at a point above the larynx, with or without accompanying vocal fold vibrations.

Despite the fact that the basis of all normal vowels is the glottal tone, we are all capable of distinguishing a large number of vowel qualities. Yet the glottal vibrations in the case of [ɑ:] are not very different from those for [i:], when both vowels are said with the same pitch. The modifications in quality which we perceive are due to the action of the supraglottal resonators which we have previously described. To understand this action, it is necessary to consider a little more closely the nature of the glottal vibrations.

It has already been mentioned that the glottal tone is the result of a complex, but mainly regular, vibratory motion. In fact, the vocal folds vibrate in such a way as to produce, in addition to a basic vibration over their whole length (the FUNDAMENTAL FREQUENCY), a number of overtones or HARMONICS having frequencies which are simple multiples of the fundamental or first harmonic. Thus, if there is a fundamental frequency of vibration of 100 Hz, the upper harmonics will be of the order of 200, 300, 400, etc. Hz. Indeed, there may be no energy at the fundamental frequency, but merely the harmonics of higher frequency such as 200, 300, 400 Hz. Nevertheless, we still perceive a pitch which is appropriate to a fundamental frequency of 100 Hz; i.e. the fundamental frequency is the highest common factor of all the frequencies present, whether or not it is present itself.

The number and strength of the component frequencies of this complex glottal tone will differ from one individual to another, and this accounts at least in part for the differences of voice quality by which we are able to recognize speakers. But we can all modify the glottal tone so as to produce at will vowels as different as [i:] and [ɑ:], so that, despite our divergences of voice quality, we can convey the distinction between two words such as *key* and *car*. This variation of quality, or timbre, of the glottal tone is achieved by the shapes which we give the resonators above the larynx—the pharynx, mouth, and nasal cavity. These chambers are capable of assuming an infinite number of shapes, each of which will have a characteristic

vibrating resonance of its own. Those harmonics of the glottal tone which coincide with the chamber's own resonance are very considerably amplified. Thus, certain bands of strongly reinforced harmonics are characteristic of a particular arrangement of the resonating chambers which produces, for instance, a certain vowel sound. Moreover, these bands of frequencies will be reinforced whatever the fundamental frequency. In other words, whatever the pitch on which we say, for instance, the vowel [ɑ:], the shaping of the resonators and their resonances will be very much the same, so that it is still possible, except on extremely high or low pitches, to recognize the quality intended. It is found that, for male speakers, the vowel [ɑ:] has one such characteristic band of strong components in the region of 700 Hz and another at about 1,100 Hz. The vowel [i:] has, for female speakers, bands of energy at about 320 and 2,700 Hz.

3.2 *The Acoustic Spectrum*

This complex range of frequencies of varying intensity which go to make up the quality of a sound is known as the **ACOUSTIC SPECTRUM**, and those bands of energy which are characteristic of a particular sound are known as the **SOUND'S FORMANTS**. Thus, formants of [ɑ:] are said to occur, for male speakers, in the region 700 and 1,100 Hz.

Such complex waveforms can be analysed and displayed as a **SPECTROGRAM**. Originally this display required a special instrument, a spectrograph, but nowadays it is generally done by computer. The spectrogram consists of a three-dimensional display: frequency is shown on the vertical axis, time on the horizontal axis, and the energy at any frequency level either by the density of blackness in a black and white display, or by colours in a colour display. Thus the concentrations of energy at particular frequency bands (the formants) stand out very clearly. Fig. 3 shows spectrograms of the vowels /ɑ:/ and /i:/ as said by a male speaker of British English. Fig. 3 also shows, in the spectrogram of *Manchester music shops*, the extent to which utterances are not neatly segmented into a succession of sounds but that, on the contrary, considerable overlap is involved. Such spectrographic analysis provides a great deal of acoustic information in a convenient form. Nevertheless, much of the information given is, in fact, irrelevant to our understanding of speech, and the phonetician is obliged to establish by other methods the elements of the spectrum which are essential to speech communication.

For instance, two, or at the most three, formants appear to be sufficient for the correct identification of vowels. As far as the English vowels are concerned, the first three formants are all included in the frequency range 0–4,000 Hz, so that the spectrum above 4,000 Hz would appear to be largely irrelevant to the recognition of our vowels. It is true that on a telephone system, which may have a frequency range of about 300–3,000 Hz, we find little difficulty in identifying the sound patterns used by a speaker and are even able to recognize voice qualities. Indeed, when we are dealing with a complete utterance in a given context, where there is a multiplicity of cues to help our understanding, a high degree of intelligibility may be retained even when there are no frequencies above 1,500 Hz.

As one would suspect, there appear to be certain relationships between the formants of vowels and the cavities of the vocal tract (i.e. the shapes taken on by

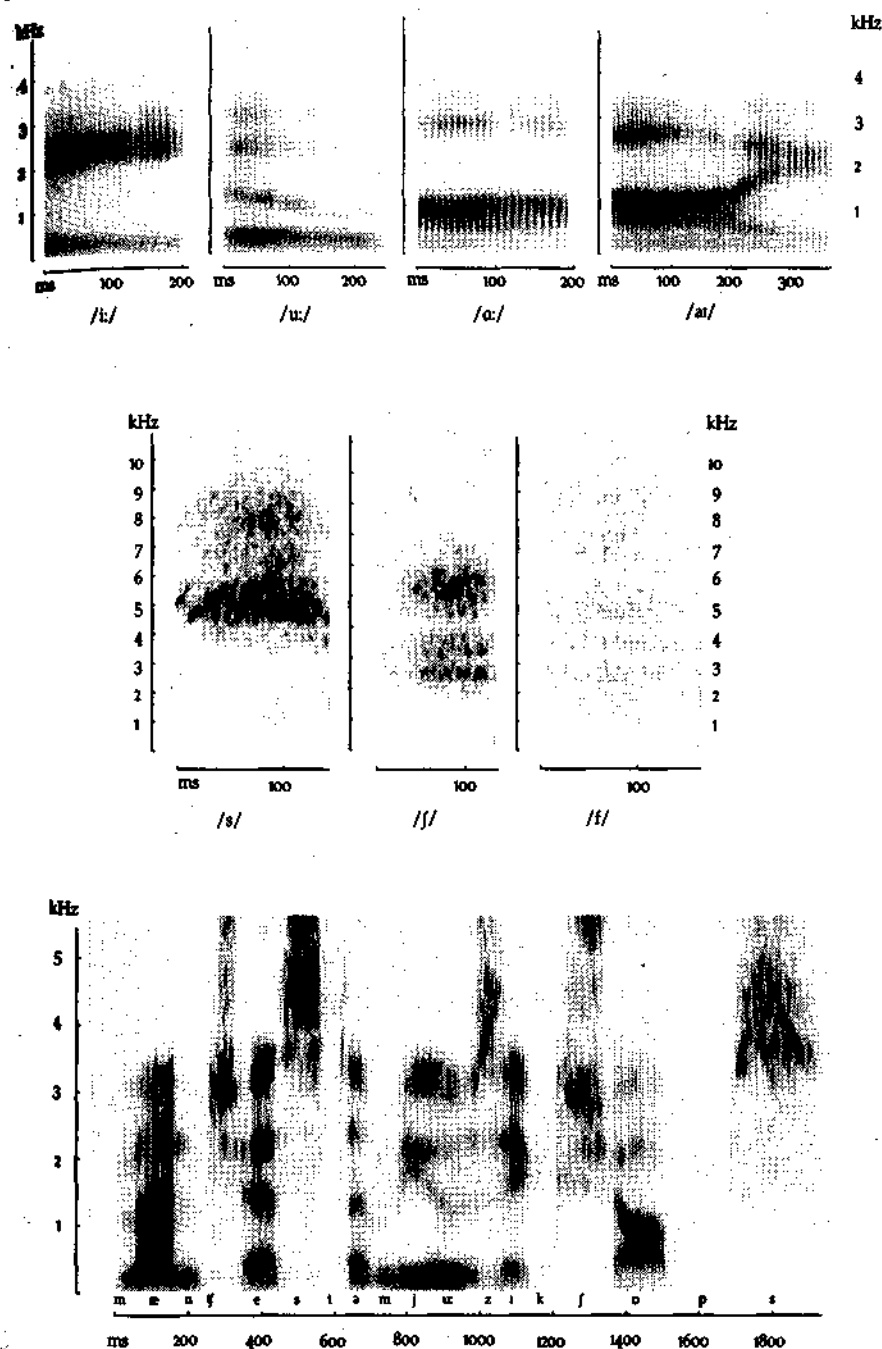


Fig. 3. Spectrograms of /i:,u:,ɑ:,a:/, /s,ʃ,f/, and *Manchester music shops*.

the resonators, notably the relation of the oral and pharyngeal cavities). Thus the first formant appears to be low when the tongue is high in the mouth: e.g. [i:] and [u:], having high tongue positions, have first formants of the order of 280–320 Hz, whereas [ɑ:] and [ɒ] have their first formants in the region 600–800 Hz, their tongue positions being relatively low. On the other hand, the second formant seems to be inversely related to the length of the front cavity: thus [i:], where the tongue is raised high in the front of the mouth, has a second formant around 2,200–2,700 Hz, whereas [u:], where the tongue is raised at the back of the mouth and lips are rounded, has a relatively low second formant around 1,100–1,400 Hz.

It is also confirmed from spectrographic analysis that a diphthong, such as that in *my*, is indeed a glide between two vowel elements (reflecting a perceptible articulatory movement), since the formants bend from those positions typical of one vowel to those characteristic of another (see Fig. 3).

For many consonant articulations (e.g. the initial sounds in *pin*, *tin*, *kin*, *thin*, *fin*, *sin*, *shin*, in which the glottal vibrations play no part) there is an essential noise component, deriving from an obstruction or constriction within the mouth, approximately within the range 2,000–8,000 Hz (see Fig. 3). This noise component is also present in analogous articulations in which vocal fold excitation is present, as in the final sounds of *ruse* and *rouge*, where we are dealing with sounds which consist of a combination of glottal tone and noise. Relevant acoustic data concerning both vowel and consonant articulations will be given in the sections dealing with individual English sounds (Chapters 8 and 9).

Spectrographic analysis also reveals the way in which there tends, on the acoustic level, to be a merging of features of units which, linguistically, we treat separately. Thus, our discrimination of [f] and [θ] sounds would appear to depend not only on the frequency and duration of the noise component but also upon a characteristic bending of the formants of the adjacent vowel. Indeed, in the case of such consonants as [p,t,k], which involve a complete obstruction of the airstream and whose release is characterized acoustically by a relatively brief burst of noise, the vowel TRANSITION between the noise and the steady state of the vowel appears to be of prime importance for our recognition of the consonant.

3.2.1 Fundamental Frequency: Pitch

Our perception of the pitch of a speech sound depends directly upon the frequency of vibration of the vocal folds. Thus we are normally conscious of the pitch caused by the voiced sounds, especially vowels; pitch judgements made on voiceless or whispered sounds, without the glottal tone, are limited in comparison with those made on voiced sounds, and are induced mainly by variations of intensity or by the dominance of certain harmonics brought about by the dispositions of the resonating cavities.

The higher the glottal fundamental frequency, the higher our impression of pitch. A male voice may have an average pitch level of about 120 Hz and a female voice a level in the region of 220 Hz.¹ The pitch level of voices, however, will vary a great deal between individuals and also within the speech of one speaker, the total range of one speaking voice being liable to have a range as extensive as 80–350 Hz. Yet

¹ Fant (1956).

our perception of frequency extends further than the limits of glottal fundamental frequency, since our recognition of quality depends upon frequencies of a much higher order. In fact, the human ear perceives frequencies from as low as 16 Hz to about 20,000 Hz and in some cases even higher. As one becomes older, this upper limit may fall considerably, so that at the age of fifty it may extend no higher than about 10,000 Hz. As we have seen, such a reduced range is no impediment to perfect understanding of speech, since a high percentage of acoustic cues for speech recognition fall within the range 0–4,000 Hz.

Our perception of pitch is not, however, solely dependent upon fundamental frequency. Variations of intensity on the same frequency may induce impressions of a change of pitch; and conversely, tones of very high or very low frequency, if they are to be audible at all, require greater intensity than those in a middle range of frequencies.

Instrumental measurement of fundamental frequency based on signals received through a microphone employs two general methods. The first is to count the number of times that a particular pattern is repeated within a selected segment of a waveform such as that provided on an oscillogram. The second is to track the progress of the fundamental frequency on a spectral display like that provided on a spectrogram, or, alternatively, to track the progress of a particular harmonic and divide by the relevant number. Nowadays various computer programs are available which average the results from a range of measurements based on the two general methods noted above. But even with such sophisticated programs there are still likely to be the occasional mistakes like octave jumps (the difference between two harmonics representing an octave perceptually).

A third method of fundamental frequency extraction involves direct measurement of the vibration of the vocal folds either by glottal illumination or by electroglottography. The best known technique in the latter class involves using a LARYNGOGRAPH.² Electrodes are attached to the outside of the throat, and the varying electrical impedance is monitored and projected onto a visual display. The signal generated by the variation in impedance can also be stored, enabling this technique to be used outside the laboratory.

Measures of fundamental frequency do not always correspond to our auditory perception of pitch. Different segments affect the fundamental frequency in different ways: for example, other things being equal, an [i] will have a higher fundamental frequency than an [a] and a [p] will produce a higher frequency on a following vowel than a [b]. What is more, many slight changes in frequency will be undetectable by the ear. As in many other cases of instrumental measurement, we still have to use our auditory perception to interpret what instruments tell us.

3.2.2 Intensity: Loudness

Our sensation of the relative loudness of sounds may depend on several factors, e.g. a sound or syllable may appear to stand out from its neighbours—be 'louder'—because a marked pitch change is associated with it or because it is longer than its neighbours. It is better to use a term such as PROMINENCE to cover these general listener-impressions of variations in the perceptibility of sounds. More strictly, what

² Abberton and Fourcin (1984).

is 'loudness' at the receiving end should be related to INTENSITY at the production stage, which in turn is related to the size of AMPLITUDE of the vibration. An increase in amplitude of vibration, with its resultant impression of greater loudness, is brought about by an increase in air pressure from the lungs. As we shall see (§10.2), this greater intensity is not in itself usually the most important factor in rendering a sound prominent in English. Moreover, all other things being equal, some sounds appear by their nature to be more prominent or sonorous than others, e.g. the vowel in *barn* has more carrying power than that in *bean*, and vowels generally are more powerful than consonants.

The judgements we make concerning loudness are not as fine as those made for either quality or pitch. We may judge which of two sounds is the louder, but we find it difficult to express the extent of the difference. Indeed, in terms of our linguistic system, we need perceive and interpret only gross differences of loudness, despite the fact that when we judge quality we are, in recognizing the formant structure of a sound, reacting to characteristic regions of strong intensity in the spectrum.

3.2.3 Duration: Length

In addition to affording different auditory impressions of quality, pitch, and loudness, sounds may appear to a listener to be of different length. Clearly, whenever it is possible to establish the boundaries of sounds or syllables, it will be possible to measure their duration by means of such traces as are provided by oscillograms or spectrograms. Such delimitation of units, in both the articulatory and acoustic sense, may be difficult, as we shall see when we deal with the segmentation of the utterance. But, even when it can be done, variations of duration in acoustic terms may not correspond to our linguistic judgements of length. We shall, for instance, refer later to the 'long' vowels of English such as those of *bean* and *barn*, as compared with the 'short' vowel in *bin*. But, in making such statements, we shall not be referring to absolute duration values, since the duration of all vowels will vary considerably from utterance to utterance, according to factors like whether the utterance is spoken fast or slowly, whether the syllable containing the vowel is accented or not, and whether the vowel is followed by a voiced or voiceless consonant. In the English system, however, we know that no more than two degrees of length are ever linguistically significant, and all absolute durations will be interpreted in terms of this relationship. This distinction between measurable duration and linguistic length provides another example of the way in which our linguistic sense interprets from the acoustic material only that which is significant.

The sounds composing any utterance will have varying durations, and we will have the impression that some syllables are longer than others. Such variations of length within the utterance constitute one manifestation of the rhythmic delivery which is characteristic of English and so is fundamentally different from the flow of other languages, such as French, where syllables tend to be of much more even length.

As already mentioned, the absolute duration of sounds or syllables will depend, among other things, upon the speed of utterance. An average rate of delivery might contain anything from about 6 to 20 sounds per second, but lower and much higher

speeds are frequently used without loss of intelligibility. The time required for the recognition of a sound will depend upon the nature of the sound and the pitch, vowels and consonants differing considerably in this respect, but it seems that a vowel lasting only about 4 msec may have a good chance of being recognized.

3.2.4 'Stress'

We have purposely avoided the use of the word 'stress' in this chapter because this word has been used in different and ambiguous ways in phonetics and linguistics. It has sometimes been used as simply equivalent to loudness, sometimes as meaning 'made prominent by means other than pitch' (i.e. by loudness or length), and sometimes as referring just to syllables in words in the lexicon and meaning something like 'having the potential for accent on utterances'. Throughout this book we will avoid use of the term 'stress' altogether, using prominence as the general term referring to segments or syllables, SONORITY as the particular term referring to the carrying power of individual sounds, and ACCENT as referring to those syllables which stand out above others, either in individual words or in longer utterances.

3.3 Hearing

Our hearing mechanism must be thought of in two ways: the physiological mechanism which reacts to the acoustic stimuli—the varying pressures in the air which constitute sound; and the psychological activity which, at the level of the brain, selects from the gross acoustic information that which is relevant in terms of the linguistic system involved. In this way, measurably different acoustic stimuli may be interpreted as being the 'same' sound unit. As we have seen, only part of the total acoustic information seems to be necessary for the perception of particular sound values. One of the tasks which confront the phonetician is the disentangling of these relevant features from the mass of acoustic material, such as modern methods of sound analysis make available. The most fruitful technique of discovering the significant acoustic cues is that of SPEECH SYNTHESIS, controlled by listeners' judgements. After all, the sounds [ɑ:] and [s] are [ɑ] and [s] only if listeners recognize them as such. Thus, it has been established that only two formants are necessary for the recognition of vowels, because machines which generate sound of the appropriate frequency bands and intensity produce vowels which are correctly identified by listeners.

Listeners without any phonetic training can, therefore, frequently give valuable guidance by their judgements of synthetic qualities. But it is important to be aware of the limitations of such listeners, so as to be able to make a proper evaluation of their judgements. A listener's reactions are normally conditioned by his experience of handling his own language. Thus, if there are only five significant vowel units in his language, he is liable to allow a great deal more latitude in his assessment of what is the 'same' vowel sound than if he has twenty. An Englishman, for instance, having a complex vowel system and being accustomed to distinguishing such subtle distinctions as those in *sit*, *set*, *sat*, will be fairly precise in his judgement of vowel

qualities. A Spaniard, however, whose vowel system is made up of fewer significant units, is likely for this reason to be more tolerant of variation of quality. Or again, if a listener is presented with a system of synthetic vowels which is numerically the same as his own, he is able to make allowance for considerable variations of quality between his and the synthetic system and still identify the vowels correctly—by their 'place' in a system rather than by their precise quality; this is what he does when he listens to and understands his language as used by a speaker of a different dialect.

Our hearing mechanism also plays an important part in monitoring our own speech; it places a control upon our speech production which is complementary to our motor, articulatory, habits. If this feedback control is disturbed, e.g. by the imposition of an artificial delay upon our reception of our own speech, disturbance in the production of our utterance is likely to result. Those who are born deaf or who become deaf before the acquisition of speech habits are rarely able to learn normal speech; similarly, a severe hearing loss later in life is likely to lead eventually to a deterioration of speech.

4

The Description and Classification of Speech Sounds

4.1 *Phonetic Description*

We have considered briefly both the mechanism which produces speech sounds and also some of the acoustic and auditory characteristics of the sounds themselves. It is now important to formulate a method of description and classification of the sound types which occur in speech and, more particularly, in English. We have seen that a speech sound has at least three stages available for investigations—the production, transmission, and reception stages. A complete description of a sound should, therefore, include information concerning all three stages. To describe the first sound in the word *ten* merely in terms of the movements of the organs of speech is to ignore the nature of the sound which is produced and the features perceived by a listener. Nevertheless, to provide all the information in respect of all phases entails a lengthy description, much of which may be irrelevant to a particular purpose. For example, since the description of the sounds of a language has in the past been most commonly used in the teaching of the language to foreigners, the emphasis has always been laid on the articulatory event. Moreover, it is only comparatively recently that there has existed any considerable body of acoustic information concerning speech. The most convenient and brief descriptive technique continues to rely either on articulatory criteria or on auditory judgements, or on a combination of both. Thus, those sounds which are commonly known as 'consonants' are most easily described mainly in terms of their articulation, whereas 'vowel' sounds require for their description a predominance of auditory impressions.

4.2 *Vowel and Consonant*

Two types of meaning are associated with the terms 'vowel' and 'consonant'. Traditionally, consonants are those segments which, in a particular language, occur at the edges of syllables, while vowels are those which occur at the centre of syllables. So, in *red, wed, dead, lead, said*, the sounds represented by < r, w, d, l, s >

are consonants, while in *beat, bit, bet, but, bought*, the sounds represented by <a,i,e,u,ough> are vowels. This reference to the functioning of sounds in syllables in a particular language is a phonological definition. But once any attempt is made to define what sorts of sounds generally occur in these different syllable-positions, then we are moving to a phonetic definition. This type of definition might define vowels as median (air must escape over the middle of the tongue, thus excluding the lateral [l]), oral (air must escape through the mouth, thus excluding nasals like [n]), frictionless (thus excluding fricatives like [s]), and continuant (thus excluding plosives like [p]); all sounds excluded from this definition would be consonants. But difficulties arise in English with this definition (and with others of this sort) because English /j,w,r/, which are consonants phonologically (functioning at the edges of syllables) are vowels phonetically. Because of this, these sounds are often called semi-vowels. The reverse type of difficulty is encountered in words like *sudden* and *little*, where the final consonants /n/ and /l/ form syllables on their own and hence must be the centre of such syllables even though they are phonetically consonants, and even though /n/ and /l/ more frequently occur at the edges of syllables, as in *net* and *let*. When occurring in words like *sudden* and *little*, nasals and laterals are called syllabic consonants.

In this chapter we will be describing and classifying speech sounds phonetically (in the next chapter we return to the phonological definitions). We shall find that consonants can be voiced or voiceless, and are most easily described wholly in articulatory terms, since we can generally feel the contacts and movements involved. Vowels, on the other hand, are voiced, and, depending as they do on subtle adjustments of the body of the tongue, are more easily described in terms of auditory relationships.

4.3 Consonants

We have seen, in the preceding chapters, that the production of a speech sound may involve the action of a source of energy, a vibrator, and the movement of certain supraglottal organs. In the case of consonantal articulations, a description must provide answers to the following questions:

- (1) Is the airstream set in motion by the lungs or by some other means? (pulmonic or non-pulmonic)
- (2) Is the airstream forced outwards or sucked inwards? (egressive or ingressive)
- (3) Do the vocal folds vibrate or not? (voiced or voiceless)
- (4) Is the soft palate raised, directing the airstream wholly through the mouth, or lowered, allowing the passage of air through the nose? (oral, or nasal or nasalized)
- (5) At what point or points and between what organs does the closure or narrowing take place? (place of articulation)
- (6) What is the type of closure or narrowing at the point of articulation? (manner of articulation)

In the case of the sound [z], occurring medially in the word *easy*, the following answers would be given:

- (1) pulmonic
- (2) egressive

- (3) voiced
- (4) oral
- (5) tongue tip-alveolar ridge
- (6) fricative

These answers provide a concise phonetic label for the sound; a more detailed description would include additional information concerning, for instance, the shape of the remainder of the tongue, the relative position of the jaws, and the lip position.

4.3.1 Egressive Pulmonic Consonants

Most speech sounds are made with egressive lung air. Virtually all English sounds are so made, the exception being [p,t,k], which in some dialects become ejectives (see §4.3.9 below).

4.3.2 Voicing

At any place of articulation, a consonantal articulation may be voiceless or voiced.

4.3.3 Place of Articulation

The chief points of articulation are the following:

BILABIAL. The two lips are the primary articulators, e.g. [p,b,m].

LABIODENTAL. The lower lip articulates with the upper teeth, e.g. [f,v].

DENTAL. The tongue tip and rims articulate with the upper teeth, e.g. [θ,ð], as in *think* and *then*.

ALVEOLAR. The tip or blade of the tongue articulates with the alveolar ridge, e.g. [t,d,l,n,s,z].

POST-ALVEOLAR. The tip (and rims) of the tongue articulate with the rear part of the alveolar ridge, e.g. [ɹ] as at the beginning of English *red*.

RETROFLEX. The tip of the tongue is curled back to articulate with the part of the hard palate immediately behind the alveolar ridge, e.g. [ɻ] such as is found in south-west British and American English pronunciation of *red*.

PALATO-ALVEOLAR. The blade, or the tip and blade, of the tongue articulates with the alveolar ridge and there is at the same time a raising of the front of the tongue towards the hard palate, e.g. [ʃ,ʒ,ʧ,ʤ] as in English *ship, measure, beach, edge*.¹

PALATAL. The front of the tongue articulates with the hard palate, e.g. [j] or [ç] as in *queue* [kju:] or [kçu:] or a very advanced type of [k.g] = [c,ç], as in French *quitter* or *guide*.

VELAR. The back of the tongue articulates with the soft palate, e.g. [k,g,ŋ], the last as in *sing*.

UVULAR. The back of the tongue articulates with the uvula, e.g. [ʁ] as in French *rouge*.

¹ Note that these are called post-alveolar on the chart of the International Phonetic Alphabet (Table 1).

GLOTTAL. An obstruction, or a narrowing causing friction but not vibration, between the vocal folds, e.g. [h].

In the case of some consonantal sounds, there may be a secondary place of articulation in addition to the primary. Thus, in the so-called 'dark' [ɫ], as at the end of *pull*, in addition to the partial alveolar contact, there is an essential raising of the back of the tongue towards the velum (velarization); or, again, some post-alveolar articulations of [ɹ] are accompanied by slight lip-rounding (labialization). The place of PRIMARY ARTICULATION is that of the greatest stricture, that which gives rise to the greatest obstruction to the airflow. The SECONDARY ARTICULATION exhibits a stricture of lesser rank. Where there are two coextensive strictures of equal rank, an example of DOUBLE ARTICULATION results.

4.3.4 Manner of Articulation

The obstruction made by the organs may be total, intermittent, or partial, or may merely constitute a narrowing sufficient to cause friction. The chief types of articulation, in decreasing degrees of closure, are as follows:

(1) Complete Closure

PLOSIVE. A complete closure at some point in the vocal tract, behind which the air pressure builds up and can be released explosively, e.g. [p, b, t, d, k, g, ʔ].

AFFRICATE. A complete closure at some point in the mouth, behind which the air pressure builds up; the separation of the organs is however slow compared with that of a plosive, so that friction is a characteristic second element of the sound, e.g. [tʃ, dʒ].

NASAL. A complete closure at some point in the mouth but, the soft palate being lowered, the air escapes through the nose. These sounds are continuants and, in the voiced form, have no noise component; they are, to this extent, vowel-like, e.g. [m, n, ŋ].

(2) Intermittent Closure

TRILL (OR ROLL). A series of rapid intermittent closures made by a flexible organ on a firmer surface, e.g. [r], where the tongue tip trills against the alveolar ridge as in Spanish *perro*, or [ʀ] where the uvula trills against the back of tongue, as in a stage pronunciation of French *rouge*.

TAP. A single tap made by a flexible organ on a firmer surface, e.g. [ɾ] where the tongue tip taps once against the teeth ridge, as in many Scottish pronunciations of English /r/.

(3) Partial Closure

LATERAL. A partial (but firm) closure is made at some point in the mouth, the airstream being allowed to escape on one or both sides of the contact. These sounds may be continuant and frictionless and therefore vowel-like (i.e. approximants like the sounds in (5) below), as in [l, ɫ], as pronounced in southern British *little* [lɪɫ], or they may be accompanied by a little friction [l̪] as in *fling* or by considerable friction [ɫ] as in *please*.

(4) Narrowing

FRICATIVE. Two organs approximate to such an extent that the airstream passes between them with friction, e.g. [f, v, θ, ð, s, z, ʃ, ʒ, ç, x, h]. In the bilabial region, a distinction is to be made between those purely bilabial such as [ɸ, β], where the friction occurs between spread lips, and a labial-velar sound like [ɱ], where the friction occurs between rounded lips and is accompanied by a characteristic modification of the mouth cavity brought about by the raising of the back of the tongue towards the velum. [ç] occurs at the beginning of *huge*, [x] and [ɱ] in Scottish pronunciations of *loch* and *which*, and [β] in Spanish *haber*.

(5) Narrowing without Friction

APPROXIMANT (OR FRICTIONLESS CONTINUANT). A narrowing is made in the mouth but the narrowing is not quite sufficient to cause friction. In being frictionless and continuant, approximants are vowel-like; however, they function phonologically as consonants, i.e. they appear at the edges of syllables. They also differ phonetically from such sounds functioning as vowels in either of two ways. Firstly, the articulation may not involve the body of the tongue, e.g. post-alveolar [ɹ] and labiodental [v], the former the usual pronunciation in RP at the beginning of *red*, the latter a speech-defective pronunciation of the same sound. Secondly, where they do involve the body of the tongue, the articulations represent only brief glides to a following vowel: thus [j] in *yet* is a glide starting from the [i] region and [w] in *wet* is a glide starting from the [u] region.

4.3.5 Obstruents and Sonorants

It is sometimes found useful to classify categories of sounds according to their noise component. Those in whose production the constriction impeding the airflow through the vocal tract is sufficient to cause noise are known as OBSTRUENTS. This category comprises plosives, fricatives, and affricates. SONORANTS are those voiced sounds in which there is no noise component (i.e. voiced nasals, approximants, and vowels).

4.3.6 Fortis and Lenis

A voiceless/voiced pair such as English /s,z/ are distinguished not only by the presence or absence of voice but also by the degree of breath and muscular effort involved in the articulation. Those English consonants which are usually voiced tend to be articulated with relatively weak energy, whereas those which are always voiceless are relatively strong. Indeed, we shall see that in certain situations the so-called voiced consonants may have very little voicing, so that the energy of articulation becomes a significant factor.

4.3.7 Classification of Consonants

The chart of the International Phonetic Alphabet (IPA) (see Table 1) shows manner of articulation on the vertical axis; place of articulation on the horizontal axis; and

a pairing within each box thus created shows voiceless consonants on the left and voiced consonants on the right.

4.3.8 Ingressive Pulmonic Consonants

Consonants of this type, made as we are breathing in, sometimes occur in languages as variants of their egressive pulmonic equivalents. So we may use such sounds when we are out of breath, but have not got time to pause, either because the need for communication is pressing or because we do not wish someone else to have a chance to speak. The use of such an ingressive pulmonic airstream is, however, variable between languages, and is not especially common in English. Individual sounds may occur as speech defects. Some sounds may also occur extralinguistically, so in English a common way of expressing surprise or pain involves the energetic inspiration of air accompanied by bilabial or labiodental friction.

4.3.9 Egressive Glottalic Consonants

In the production of these sounds, known as EJECTIVE, the glottis is closed, so that lung air is contained beneath it. A closure or narrowing is made at some point above the glottis (the soft palate being raised) and the air between this point and the glottis is compressed by a general muscular constriction of the chamber and a raising of the larynx. Thus a bilabial ejective plosive sound [p'] may be made by compressing the air in this way behind the lips. However, it is not only plosives which may be ejective; affricates and fricatives commonly have this type of compression in a number of languages, e.g. [ts',t',s',x']. If the glottis is tightly closed, it follows that this type of articulation can apply only to voiceless sounds. [p',t',k'] occur sometimes in final positions in some dialects of English (e.g. in south-east Lancashire). These are not to be confused with the more common variants of final [p,t,k] which are frequently (e.g. in London English) replaced or reinforced by a glottal stop; i.e. the final sound in the word *stop* may be replaced by a glottal stop or have a glottal closure accompanying the bilabial one, but there is no compression between the glottal and bilabial closures.

4.3.10 Ingressive Glottalic Consonants

For these sounds a complete closure is made in the mouth but, instead of air pressure from the lungs being compressed behind the closure, the almost completely closed larynx is lowered so that the air in the mouth and pharyngeal cavities is rarefied. The result is that outside air is sucked in once the mouth closure is released; at the same time, there is sufficient leakage of lung air through the glottis to produce voice. It will be seen that the resulting sound is made by means of a combined airstream mechanism, namely a pulmonic airstream in combination with ingressive glottalic air. Such ingressive stops (generally voiced) are known as IMPLOSIVES and occur with bilabial [ɓ], dental or alveolar [ɗ], or velar [ɠ] mouth closures. Though such sounds occur in a number of languages, sometimes in the

Table 1. The International Phonetic Alphabet (revised to 1989)

speech of the deaf, and in types of stammering, they are not found in normal English. In some languages voiceless implosives may occur, which of course means that in these cases the larynx must be completely closed.

4.3.11 Ingressive Velaric Consonants

Another set of sounds involving an ingressive airstream is produced entirely by means of closures within the mouth cavity; normal breathing through the nose may continue quite independently if the soft palate is lowered, and may even produce accompanying nasalization. Thus, the sound made to indicate irritation or sympathy (often written as 'tut-tut') is articulated by means of a double closure, the back of the tongue against the velum and the tip, blade, and sides against the alveolar ridge and side teeth. The cavity contained within these closures is then enlarged mainly by tongue movement, so that the air is rarefied. The release of the forward closure causes the outer air to be sucked in; the release may be crisp, in which case a sound of a plosive type is heard, or relatively slow, in which case an affricated sound is produced. These sounds are known as **CLICKS**, the one referred to above being a dental click [l]. The sound made to encourage horses is a lateral click, i.e. the air is sucked in by releasing one side of the tongue [ll]. These clicks and several others occur as significant sounds in a small number of languages in Africa (e.g. Zulu) and paralinguistically in most languages (as in English).

4.4 Vowels

This category of sounds is normally made with a voiced egressive airstream, without any closure or narrowing such as would result in the noise component characteristic of many consonantal sounds; moreover, the escape of the air is characteristically accomplished in an unimpeded way over the middle line of the tongue. We are now concerned with a glottal tone modified by the action of the upper resonators of the mouth, pharyngeal, and nasal cavities. As we have seen (Chapters 2 and 3), the movable organs mainly responsible for shaping these resonators are the soft palate, lips, and tongue. A description of vowel-like sounds must, therefore, note:

- (1) the position of the soft palate—raised for oral vowels, lowered for nasalized vowels;
- (2) the kind of aperture formed by the lips—degrees of spreading or rounding;
- (3) the part of tongue which is raised and the degree of raising.

Of these three factors, only the second—the lip position—can be easily described by visual or tactile means. Our judgement of the action of the soft palate depends less on our feeling for its position than on our perception of the presence or absence of nasality in the sound produced. Again, the movements of the tongue, which so largely determine the shape of the mouth and pharyngeal cavities, may be so minute that it is impossible to assess them by any simple means; moreover, there being normally no contact of the tongue with the roof of the mouth, no help is given by any tactile sensation. A vowel description will usually, therefore, be based mainly on auditory judgements of sound relationships, together with some articulatory information, especially as regards the position of the lips. In addition, an acoustic

description can be given in terms of the disposition of the characteristic formants of the sound.

4.4.1 Difficulties of Description

The description of vowel sounds, especially by means of the written word, has always presented considerable difficulty. Certain positions and gross movements of the tongue can be felt. We are, for instance, aware that when we pronounce most vowel sounds the tongue tip lies behind the lower teeth; moreover, in comparing two such vowels as /i:/ (*key*) and /ɑ:/ (*car*) (Fig. 4), we can feel that, in the case of the former, the front of the tongue is the part which is mainly raised, whereas in the case of the latter, such raising as there is is accomplished by the back part of the tongue. Therefore, it can be stated in articulatory terms that some vowel sounds require the raising of the front of the tongue, while others are articulated with a typical 'hump' at the back; and these statements can be confirmed by means of x-ray photography. But the actual point and degree of raising is more difficult to judge. It is not, for instance, helpful to say that a certain vowel is articulated with the front part of the tongue raised to within 5mm of the hard palate. This may be a statement of fact for one person's pronunciation, but an identical sound may be produced by another speaker with a different relationship between the tongue and palate. Moreover, we would not find it easy to judge whether our tongue was at 4 or 5 mm from the palate. It is no more helpful to relate the vowel quality to a value used in a particular language, as is still so often done. A statement such as 'a vowel quality similar to that in the English word *cat*' is not precise, since the vowel in *cat* may have a wide range of values in English. The statement becomes more useful if the accent of English is specified, but even then a number of variant interpretations will always be possible.

4.4.2 Cardinal Vowels

It is clear that a finer and more independent system of description is needed, on both the auditory and articulatory levels. The most satisfactory scheme is that devised by Daniel Jones and known as the **CARDINAL VOWEL SYSTEM**. The basis of the

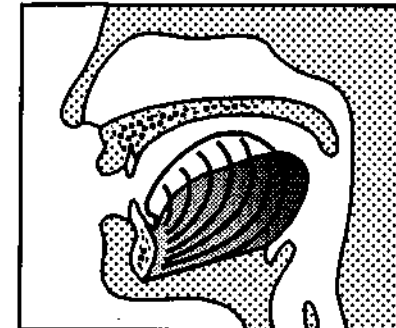


Fig. 4. Tongue positions of [i:], [ɑ:].

system is physiological, i.e. the two qualities upon which all the others were 'hinged' were produced with the tongue in certain easily felt positions: the front of the tongue raised as close as possible to the palate without friction being produced, for the Cardinal Vowel [i]; and the whole of the tongue as low as possible in the mouth, with very slight raising at the extreme back, for the Cardinal Vowel [a]. Starting from the [i] position, the front of the tongue was lowered gradually, the lips remaining spread or neutrally open and the soft palate raised. The lowering of the tongue was halted at three points at which the vowel qualities seemed, from an auditory standpoint, to be equidistant. The tongue positions of these qualities were x-rayed and were found to be more or less equidistant from a spatial point of view. The symbols [e, ε, a] were assigned to these vowel values. The same procedure was applied to vowel qualities depending on the height of the back of the tongue, thus raising the back of the tongue from the [a] position; the lips were changed progressively from a wide open shape (for [a]) to a closely rounded one and the soft palate remained raised. Again, three auditorily equidistant points were established from the lowest to the highest position; the corresponding tongue positions were photographed and the spatial relationships confirmed as for the front vowels. These values were given the symbols [ɔ, o, u]. Thus a scale of eight primary Cardinal Vowels was set up, denoted by the following numbers and symbols: 1, [i]; 2, [e]; 3, [ε]; 4, [a]; 5, [ɑ]; 6, [ɔ]; 7, [o]; 8, [u].

It is to be noticed that the front series [i, e, ε, a] and [ɑ] of the back series are pronounced with spread or open lips, whereas the remaining three members of the back series have varying degrees of lip-rounding. The combination of tongue and lip positions in the primary Cardinal Vowels are the most frequent in languages; i.e. front and open vowels are most commonly unrounded while back vowels other than in the open position are most commonly rounded. A secondary series can be obtained by reversing the lip positions, e.g. lip-rounding applied to the [i] tongue position, or lip-spreading applied to the [u] position. Such a secondary series is denoted by the following numbers and symbols: 9, [y]; 10, [ø]; 11, [œ]; 12, [œ̃]; 13, [ɒ]; 14, [ʌ]; 15, [ɹ]; 16, [ʊ].

This complete series of sixteen Cardinal Vowel values may be divided into two lip shape categories, with corresponding tongue positions:

unrounded: [i, e, ε, a, ɑ, ɹ, u].

rounded: [y, ø, œ, œ̃, ɒ, ʌ, ɹ, ʊ].

Such a scale is useful because (a) the vowel qualities are unrelated to particular values in languages, though many may occur in various languages, and (b) the set is recorded, so that reference may always be made to a standard, invariable scale.² Thus a vowel quality can be described as being, for instance, similar to that of Cardinal 2 ([e]), or another as being a type half-way between Cardinal 6 ([ɔ]) and Cardinal 7 ([o]), but somewhat centralized. Diacritics are available in the IPA alphabet to show modifications of Cardinal values, e.g. a subscript ₁ to mean more open, a subscript ₂ meaning closer, and raised dots [˘] to mean centralized. The last example given above might in this way be symbolized as [ɔ̃₁] or [ɔ̃₂].

It is, moreover, possible to give a visual representation of these vowel relation-

ships on a chart which is based on the Cardinal Vowel tongue positions. The simplified diagram shown in Fig. 5 is obtained by plotting the highest point of tongue-raising for each of the primary Cardinal Vowels and joining the points together. The internal triangle, corresponding to the region of central or [ə]-type vowel sounds, is made by dividing the top line into three approximately equal sections and drawing lines parallel to the two sides, so that they meet near the base of the figure. On such a figure, the sound symbolized by [ɜ̃] or [ø̃] may have its relationship to the Cardinal scale shown visually (see the black circle on Fig. 5).

It must be understood that this diagram is a highly conventionalized one which shows, above all, quality relationships. Some attempt is, however, made to relate the shape of the figure to actual tongue positions: thus the range of movement is greater at the top of the figure, and the tongue-raising of front vowels becomes more retracted as the tongue position lowers. Nevertheless, it has been shown that it is possible to articulate vowel qualities without the tongue and lip positions which this diagram seems to postulate as necessary. It is, for instance, possible to produce a sound of the Cardinal 7 ([o]) type without the lip-tongue relationship suggested. But, on the whole, it may be assumed that a certain auditorily identified vowel quality will be produced by an articulation of the kind presupposed by the Cardinal Vowel diagram. Moreover, it is a remarkable fact that the auditory judgements as to vowel relationships made by Daniel Jones have been largely supported by recent acoustic analysis; in fact, a chart based on an acoustic analysis of Cardinal Vowel qualities corresponds very well with the traditional Cardinal Vowel figure.

4.4.3 Nasality

Besides the information concerning lip and tongue positions which the chart and symbolization denote, a vowel description must also indicate whether the vowel is purely oral or whether it is nasalized. The sixteen Cardinal Vowels mentioned may all be transformed into their nasalized counterparts if the soft palate is lowered. It is unusual, however, to find such an extensive series of nasalized vowels, since it is unusual (though not unknown) for languages to make such fine, significant,

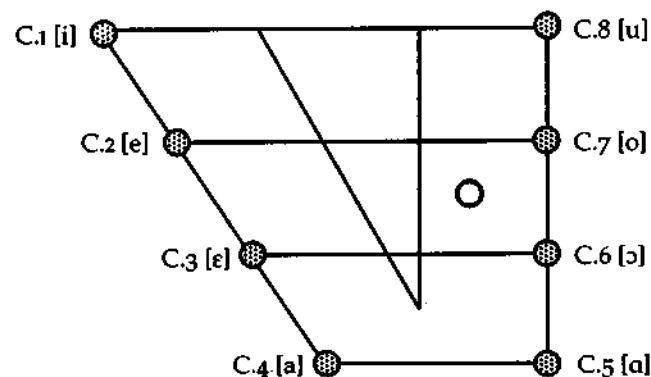


Fig. 5. The primary Cardinal Vowels; the area symbolized by [ɜ̃] or [ø̃] is shown as a circle.

² Copies of the original recording of the Cardinal vowels by Daniel Jones are available from the Phonetics Laboratory, Department of Linguistics, University of Manchester, Manchester M13 9PL.

distinctions of nasalized qualities as are common in the case of the purely oral values.

4.4.4 Relatively Pure Vowels v. Gliding Vowels

It is clearly not possible for the quality of a vowel to remain absolutely constant (or, in other words, for the organs of speech to function for any length of time in an unchanging way). Nevertheless, we may distinguish between those vowels which are relatively pure (or unchanging), such as the vowel in *learn*, and those which have a considerable and voluntary glide, such as the gliding vowel in *line*. The so-called pure vowels will be marked on the diagram as a dot, showing the highest point of the tongue, or, better, as a ring, since it would be inadvisable to attempt to be over-precise in the matter of these auditory judgements; the gliding (or diphthongal) vowel sound will be shown as an arrow, which indicates the quality of the starting-point and the direction in which the quality change is made (corresponding to a movement of the tongue). Fig. 6 shows the way in which the vowels of *learn* and *line* will be marked.

We are now in a position to give a practical and comprehensive description of a vowel sound, partly in articulatory terms, partly in auditory terms. The vowel which we have symbolized above as [ɜ̃] or [ɝ̃] might be described in this way: 'A vowel quality between Cardinal Vowels nos. 6 and 7, but having a somewhat centralized value; the lips are fairly closely rounded; and the soft palate is raised'. Such a written description will have a meaning in terms of sound for anyone who is familiar with the Cardinal vowel scale. There may, of course, be other features of the sound to mention, e.g. a breathy or creaky voice quality.

4.4.5 Articulatory Classification of Vowels

Although precise descriptions of vowels are better done auditorily, nevertheless it is convenient to have available a rough scheme of articulatory classification. Such a scheme is represented by the vowel diagram on the chart of the International Phonetic Alphabet (IPA) as shown in Table 1. It will immediately be noticed that

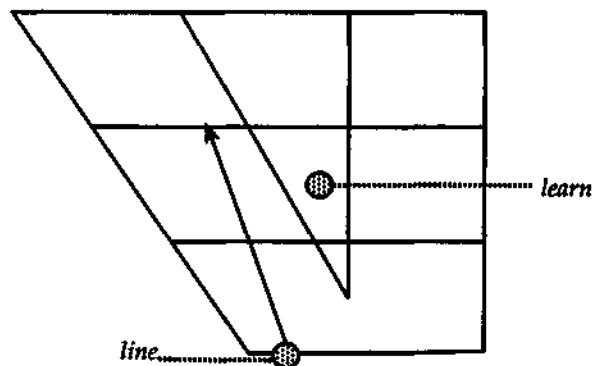


Fig. 6. The vowels of *learn* and *line*.

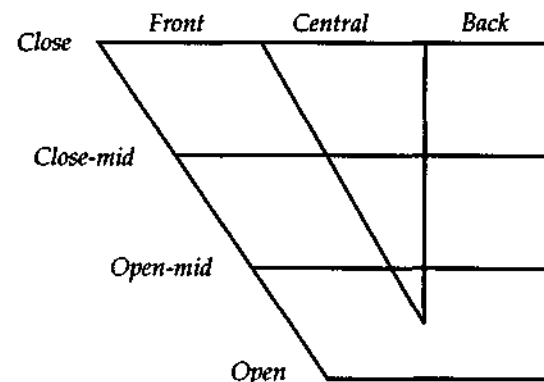


Fig. 7. Classification labels combined with the Cardinal Vowel diagram.

this is of similar shape to the Cardinal Vowel diagram. Labels are provided to distinguish between front, central, and back, and between four degrees of opening: close, close-mid, open-mid, and open (see Fig. 7). At each intersection point on the periphery of the diagram on the IPA chart (Table 1) two symbols are supplied; these symbols are the same as those used for the Cardinal Vowels. However, on this chart the unrounded vowel is always the first of the pair and the rounded the second; this means that we cannot say that the first corresponds to the primary Cardinal and the second to the secondary Cardinal. (It will be remembered that primary Cardinals involve the most frequent lip positions, back vowels being more usually rounded.) The IPA diagram also supplies us with a number of additional symbols for vowels in certain positions, [ɪ, æ, i, ə, e] being used for unrounded vowels and [ʊ, u, o] for rounded vowels. Of course, this chart does not show nasalized vowels.

5

Sounds in Language

5.1 Speech Sounds and Linguistic Units

We have now considered a method of describing and classifying the sounds capable of being produced by the speech organs. Speech is, however, a manifestation of language, and spoken language is normally a continuum of sound. A speech sound, produced in isolation and without the meaningfulness imposed by a linguistic system, may be described in purely phonetic terms; but any purely phonetic approach to the sounds of language encounters considerable difficulties. Two initial problems concern, first, the identification and delimitation of the sound unit (or segment) to be described and, secondly, the way in which different sounds are treated, for the purpose of linguistic analysis, as if they were the same.

As we have seen, in any investigation of speech, it is at the physiological and acoustic levels that most information is available to us. Yet, especially on the articulatory level, as is revealed by moving x-ray photography, any utterance consists of apparently continuous movements by a very large number of organs; it is well-nigh impossible to say, simply from an x-ray film of the speech organs at work, how many speech sounds have been uttered. A display of acoustic information is easier to handle (see Fig. 3), but even here it is not always possible, because of the way in which many sounds merge into one another, to delimit exactly the beginning and end of sound segments. Moreover, even if it were possible to identify the main characteristics of certain sounds without being sure of their limits, it would not follow that the phonetic statement we might accordingly make concerning the sequence of sound segments would be a useful one in terms of the language which we were investigating. Thus, the word *tot* is frequently pronounced in the London region in such a way that it is possible to identify five sound segments: [t], [s], [h], [v], [t]. Yet much of this phonetic reality may be discarded as irrelevant when it is a question of the structure of the word *tot* in terms of the sound system of English. Indeed, the speaker himself will probably feel that the utterance *tot* consists of only three 'sounds' (and not only because of the influence of the spelling), such a judgement on his part being a highly sophisticated one which results from his experience in hearing and speaking English. In other words, the [s] and [h] segments

are to be treated as part of the phonological, or linguistic, unit /t/.¹ The phonetic sequence [tsh] does not, in an initial position in this type of English, consist of three meaningful units; in other languages, on the other hand, such a sequence might well constitute three linguistic units as well as three phonetic segments.

This same example illustrates how different sounds may count, in respect of their function in a language, as the same linguistic unit. In such a pronunciation of *tot* as is noted above, the first realization of /t/ might be described as consisting of:

- (1) a voiceless stop made by the tongue tip and rims against the alveolar ridge and side teeth;
- (2) a slow release of the compressed air, so that friction is heard—[s];
- (3) the complete disengagement of the tongue from the roof of the mouth, so that no friction is caused in the mouth; but an interval before the beginning of the next sound, during which there is friction in the glottis (and voiceless resonance in the supraglottal cavities)—[h].

The second manifestation of /t/, on the other hand, might have an articulation which could be described phonetically as follows:

- (1) an alveolar stop made as before, but with a simultaneous stop made in the glottis;
- (2) the glottal closure is released, but the oral stop is retained slightly longer, during which time the air escapes through the nasal passage, the soft palate being lowered.

The first [t] might be briefly described as a voiceless alveolar plosive, released with affrication and aspiration; the second as an unexploded voiceless alveolar plosive made with a simultaneous glottal stop. These two different articulations, with the resultant difference of sound, nevertheless function as the same linguistic unit, the first sound occurring predictably under strong accent initially in a syllable and the second being a typical manifestation of the unit in a final position. Such an abstract linguistic unit, which will include sounds of different types, is called a **PHONEME**; the different phonetic realizations of a phoneme are known as its **ALLOPHONES**.

5.2 The Linguistic Hierarchy

It is clear, as we hinted in Chapter 1, that speech and language require in their analysis different types of unit. An utterance, on the concrete speech level, will consist of the continuous physiological activity which results in a continuum of sound; the largest unit will, therefore, be the span of sound occurring between two silences. Within this unit of varying extent it may be possible to find smaller segments. It is, however, from the abstract, linguistic, level of analysis that we receive guidance as to how the utterance may be usefully segmented in the case of any particular language. We might find, for instance, that an utterance such as 'The boys ran quickly away and were soon out of sight' is spoken without a pause or interruption for breath; it might be said to constitute a single breath group on the articulatory level. But, on the linguistic level, we know that this utterance is capable of being analysed as a sentence consisting of two **CLAUSES**. Moreover, certain

¹ It is customary to distinguish sound segments from linguistic sound units (phonemes) by using [] to enclose the former and / / to enclose the latter.

extensive sequences occurring within the utterance might be meaningfully replaced by other sound sequences, e.g. *boys* might be replaced by *dogs*, *ran* by *walked*, *quickly* by *slowly*, etc. These replaceable sound sequences are able to stand by themselves and are called **WORDS**. In written forms of language, it usually happens that words are separated from each other by spaces, this being a sophisticated convention which is not reflected in speech. (We shall see, however, in Chapter 11, that words may retain at their boundaries certain characteristics in connected speech, so that their presence and span is signalled on a phonetic as well as a linguistic level.) Yet there are meaningful units smaller than the word. The word *boys* may be divided into *boy* and *s* ([z]), where the presence or absence of [z] indicates the plural or singular form; *quickly* may be said to consist of *quick* and the adverbial suffix *-ly*. These are smaller sound sequences which may be interchanged meaningfully, but which may or may not be capable of standing by themselves. These smaller units, known as **MORPHEMES**, may correspond with words, e.g. *boy*, in which case they may stand alone, or they may not normally occur other than in association with a word. There is, however, a yet lower level at which meaningful commutation is possible. The word *ran* is also a morpheme; but if, instead of saying [ræn] we say [rʌn], we have, by changing an element on a lower level than the morpheme, changed the meaning and function of the word. This basic linguistic element, beyond which it is not necessary to go for practical purposes, is what we have already referred to as a **PHONEME**. A phoneme may, therefore, be thought of as the smallest contrastive linguistic unit which may bring about a change of meaning.

5.3 Phonemes

It is possible to establish the phonemes of a language by means of a process of commutation or the discovery of **MINIMAL PAIRS**, i.e. pairs of words which are different in respect of only one sound segment. The series of words *pin*, *bin*, *tin*, *din*, *kin*, *chin*, *gin*, *fin*, *thin*, *sin*, *shin*, *win* supplies us with twelve words which are distinguished simply by a change in the first (consonantal) element of the sound sequence. These elements, or phonemes, are said to be in **CONTRAST** or **OPPOSITION**; we may symbolize them as /p,b,t,d,k,ŋ,ç,f,θ,s,ʃ,w/. But other sound sequences will show other consonantal oppositions, e.g.

- (1) *tame*, *dame*, *game*, *lame*, *mame*, *name*, adding /g,l,m,n/ to our inventory;
- (2) *pot*, *tot*, *cot*, *lot*, *yacht*, *hot*, *rot*, adding /j,h,r/;
- (3) *pie*, *tie*, *buy*, *thigh*, *thy*, *vie*, adding /ð,v/;
- (4) *two*, *do*, *who*, *woo*, *zoo*, adding /z/.

Such comparative procedures reveal twenty-two consonantal phonemes capable of contrastive function initially in a word.

It is not sufficient, however, to consider merely one position in the word. Possibilities of phonemic opposition have to be investigated in medial and final positions as well as in the initial. If this is done in English, we discover in medial positions another consonantal phoneme, /z/, cf. the word oppositions *letter*, *leather*, *leisure* or *seater*, *seeker*, *Caesar*, *seizure*. This phoneme /z/ does not occur in initial positions and is rare (e.g. in *rouge*) in final positions. Moreover, in final positions, we do not find /h/ or /r/, and it is questionable whether we should

consider /w,j/ as separate, final, contrastive units (see §8.2). We do, however, find one more phoneme that is common in medial and final positions but unknown initially, viz. /ŋ/; cf. *simmer*, *sinner*, *singer* or *some*, *son*, *sung*.

Such an analysis of the consonantal phonemes of English will give us a total of twenty-four phonemes, of which four (/h,r,z,ŋ/) are of **restricted occurrence**—or six, if /w,j/ are not admitted finally. Similar procedures may be used to establish the vowel phonemes of English (see Chapter 8).

The final inventories of vowel and consonant phonemes will constitute a statement of the total oppositions in all positions in the word or syllable; when any particular place in the word or syllable is taken into consideration, the number of terms in the series of oppositions is likely to be more restricted.

5.3.1 Diversity of Phonemic Solutions

It is important to emphasize the fact that it is frequently possible to make several different statements of the phonemic structure of a language, all of which may be equally valid from a logical standpoint. The solution chosen will be the one which is most convenient as regards the use to which the phonemic analysis is to be put. Thus one solution might be appropriate when it is a question of teaching a language to a particular group of foreigners, when similarities and differences between two languages may need to be underlined; another solution might be appropriate if it is a question of using the phonemic analysis as a basis for an orthography, when sociolinguistic considerations (e.g. relations with other countries having particular orthographic conventions) have to be taken into account. Even without such considerations, discrepancies in analysis frequently arise in the case of such sound combinations as affricates (e.g. [ʃ,ç,ʧ,ʤ,ʧr,ʤr]) and diphthongs (e.g. [eɪ,əʊ,ɑɪ,əʊ]), which may be treated as single phonemes or combinations of two. Such problems concerning particular English sounds will be dealt with when vowels and consonants are considered in detail.

5.3.2 Distinctive Features

Up to now we have obtained an inventory of phonemes for English which is no more than a set of relationships or oppositions. The essence of the phoneme /p/, for instance, is that it is not /t/ or /k/ or /s/, etc. This is a negative definition, which it is desirable to amplify by means of positive information of a phonetic type. Thus we may say that /p/ is, from a phonetic point of view, characteristically voiceless (compared with voiced /b/); labial (compared with the places of articulation of such sounds as /t/ or /k/); and plosive (compared with /f/). The /p/ phoneme may, therefore, be defined positively by stating the combination of distinctive features which identify it within the English phonemic system: voiceless, labial, plosive.

As originally conceived, the distinctive features of a language were stated in articulatory terms using as a basis the phonetic classification of consonants described in the previous chapter. So the distinctive features of English /p/ were voiceless, labial, and plosive. Here there are three dimensions of variation: voicing,

place, and manner. But it was conceded that the distinctive features of a language might involve more or less than three dimensions. For example, in some languages (e.g. in Tamil, a language of south India) voicing is not a distinctive feature (so changing from [p] to [b] does not bring about a change of meaning), and thus only place and manner are distinctive. In other languages we may need to state four dimensions of variation. In Hindi not only is voicing (and place and manner) distinctive but aspiration is also separately distinctive from voice; compare /kaan/ 'ear', /k^haan/ 'mine', /gaan/ 'anthem', /g^haan/ 'quantity'. Such articulatory distinctive features sometimes involve two terms (voiceless v. voiced, aspirated v. unaspirated), sometimes three (e.g. labial /p,b/ v. alveolar /t,d/ v. velar /k,g/ in English), and sometimes more.

Later developments in the theory of distinctive features have involved explaining all the contrasts of a language in terms of BINARY distinctive features and suggesting that there is a set of binary features (involving around twelve or thirteen distinctions) which will account for all languages. An apparent three-term distinction like labial v. alveolar v. velar is turned into two features with plus or minus values; using 'coronal' to mean 'made with the blade of the tongue raised above the neutral position' and 'anterior' to mean 'made in front of the hard palate', the English plosives /p,b,t,d,k,g/ are then defined as follows:

	p,b	t,d	k,g
coronal	-	+	-
anterior	+	+	-

In the most well-known set of binary distinctive features², many features are still articulatory, although some are auditory or acoustic (e.g. 'strident').

In this book we use distinctive feature analysis (of the more traditional kind which allows non-binary dimensions) where such analysis is not in doubt and where it is obviously explanatory. This means that we frequently refer to feature analysis when describing the consonants of English, but use it very little when describing the vowels, since almost all distinctive feature analysis in this area is disputed and not always helpful.

5.3.3 Allophones

No two realizations of a phoneme are the same. This is true even when the same word is repeated; thus, when the word *cat* is said twice, there are likely to be slight phonetic variations in the two realizations of the phoneme sequence /k+æ+t/. Nevertheless, the phonetic similarities between the utterances will probably be more striking than the differences. But variants of the same phoneme occurring in different words or in different positions in a word will frequently show consistent phonetic differences; such consistent variants are referred to as ALLOPHONES. We have seen (§5.1) how different the initial and final allophones of /t/ in the word *tot* may be. Or again, the [k] sounds which occur initially in the words *key* and *car* are phonetically clearly different: the first can be felt to be a forward articulation, near the hard palate, whereas the second is made further back, on the soft palate. This difference of articulation is brought about by the nature of the following vowel, [i:],

having a more advanced articulation than [ɑ:]; the allophonic variation is in this case conditioned by the context. In some varieties of English the two [l] sounds of *lull* [lʌt] show a variation of a different kind. The first [l], the so-called 'clear' [l] with a front vowel resonance, has a quality very different from that of the final 'dark' [ɫ] with a back vowel resonance. Here the difference of quality is related to the position of the phoneme in the word or syllable and is not dependent upon the phonetic context, i.e. the adjacent sounds. It is possible, therefore, to predict in a given language which allophones of a phoneme will occur in any particular context or situation: they are said to be in COMPLEMENTARY DISTRIBUTION. Statements of complementary distribution can refer to preceding or following sounds (e.g. fronted [k] before front vowels like /i:/ in *key* but retracted [k] before back vowels like /ɑ:/ in *car*); to positions in syllables (plosives are strongly aspirated when initial in accented syllables); or to positions in any grammatical unit, e.g. words (vowels may optionally be preceded by a glottal stop when word-initial) or morphemes (Cockney has a different allophone of /ɔ:/ in morpheme-medial and morpheme-final positions (cf. board [bɔəd] v. bored [bɔəd])).

Complementary distribution does not take into account those variant realizations of the same phoneme in the same situation which may constitute the difference between two utterances of the same word. When the same speaker produces noticeably different pronunciations of the word *cat* (e.g. by exploding or not exploding the final /t/), the different realizations of the phonemes are said to be in FREE VARIATION. Again, the word *very* may be pronounced [veɪ] (where the middle consonant is an approximant) or [veri] (where the middle consonant is a flap). The approximant and the flap are here in free variation. Variants in free variation are also allophones (since, like those in complementary distribution, they are not involved in changes of meaning).

It is usually the case that there is some phonetic similarity between the allophones of a phoneme: for example, both the [l] sounds discussed above, as well as the voiceless fricative variety which follows /p/ or /k/ in words such as *please* and *clean*, are lateral articulations. It sometimes happens that two sounds occur in complementary distribution, but are not treated as allophones of the same phoneme because of their total phonetic dissimilarity. This is the case of [h] and [ŋ] in English; they are never significantly opposed, since [h] occurs typically in initial positions in the syllable or word, and [ŋ] in final positions. A purely logical arrangement might include these two sounds within the same phoneme, so that *hung* might be transcribed phonemically as either /hʌŋ/ or /ŋʌŋ/; but such a solution would ignore the total lack of phonetic similarity and also the feeling of native speakers. The ordinary native speaker is, in fact, often unaware of the allophonic variations of his phonemes and will, for instance, say that the various allophones of /l/ we have discussed are the 'same' sound; [h] and [ŋ], however, he will always consider to be 'different' sounds. When he makes a statement of this kind, he is usually referring to the function of the sounds in the language system, and can thereby offer helpful, intuitive information regarding the phonemic organization of his language. In the case of a language such as English, prejudices induced by the existence of written forms have naturally to be taken into account in evaluating the native speaker's reaction.

² Chomsky and Halle (1968).

5.3.4 Neutralization

It sometimes happens that a sound may be assigned to either of two phonemes with equal validity. In English, examples of this kind are to be found in the plosive series. The contrast between English /p,t,k/ and /b,d,g/ is shown in word-initial position by pairs like *pin/bin*, *team/deem*, *come/gum*. However, following /s/ there is no such contrast. Words beginning /sp-, st-, sk-/ are not contrasted with words beginning /sb-, sd-, sg-/ , although a distinction sometimes occurs word-medially, as in *disperse/disburse* and *discussed/disgust* (suggesting a syllable division between the /s/ and the following plosive). In such circumstances we say that the contrast between /p,t,k/ and /b,d,g/, the contrast between voiceless and voiced, is neutralized following /s/ in word-initial position. Words like *spin*, *steam*, and *scar* could equally well be transcribed with /b,d,g/ as with /p,t,k/. Indeed, even though the writing system itself suggests /p,t,k/ (/k/ may be written with <k> or <ç>), the sounds which actually occur following /s/ can in some respects be considered closer to /b,d,g/, since the aspiration which generally accompanies /p,t,k/ in initial position is not present after /s/ (although vowels following /p,t,k/ generally start from a higher pitch, and vowels following /sp, st, sk/ have this higher pitch, which argues for /p,t,k/).³

Another case of neutralization concerns the allophones of /m/ and /n/ before /f/ or /v/, in words like *symphony* and *infant*. The nasal consonant in each case is likely to be [m] in rapid speech, i.e. a labiodental sound anticipating the labiodental [f]. Here again, /m/ and /n/ are not opposed, so that the sound could be allocated to either the /m/ or the /n/ phoneme. In practice, since in a slow pronunciation an [m] would tend to be used in *symphony* and an [n] in *infant*, the [m] is usually regarded as an allophone of /m/ in the one case and of /n/ in the other.

5.3.5 Phonemic Systems

Statements concerning phonemic categories and allophonic variants can be made in respect of only one variety of one language. It does not follow that, because [l] and [ɫ] are not contrastive in English and belong to the same phoneme, this is so in other languages—in some kinds of Polish [l] and [ɫ] constitute separate phonemes. Or again, although /ŋ/ is a phoneme in English, in Italian the velar nasal [ŋ] is an allophone of /n/ which occurs before /k/ and /g/. Indeed, in English, too, /ŋ/ has not always had phonemic status. Nowadays, [ŋ] might be considered an allophone of /n/ before /k/ and /g/, as in *sink* and *finger*, were it not for the fact that the /g/ in words such as *sing* was lost about 400 years ago; once this situation had arisen, a phonemic opposition existed between *sin* and *sing*. In some parts of north-west England, the situation is still the same as it was 400 years ago; e.g. not only is *sink* pronounced [sɪŋk] but *sing* is pronounced [sɪŋ], and in such dialects [ŋ] can be considered an allophone of /n/.

Thus the number of phonemes may differ as between various varieties of the same language. In present-day southern British English, the words *cat*, *half*, *cart* contain the phonemes /æ/, /ɑ:/, and /ɑ:/ respectively. But one type of Scottish

English has only one vowel phoneme for all three words, the words being phonemically /kat, haf, kart/ (the pre-consonantal /r/ being pronounced). Such a dialect of English has one phoneme less than southern British English, since the opposition *Sam/psalm* is lost. On the other hand, this smaller number of phonemes is sometimes counterbalanced by the regular opposition of the first elements of such a pair as *witch/which*, which establishes a phonemic contrast between /w/ and /m/.

It should not, however, be assumed that the phonemic systems of two dialects are different only because they have a lesser or greater number of phonemes in certain areas. The sound sequence [set], i.e. with a vowel in the region of Cardinal 3, may be a realization of *sat* in one dialect and of *set* in another; the phonemic categories commonly represented as /i,e,æ/, etc., may nevertheless be present in both dialects, the vowel system in the first dialect being somewhat more 'closed' than that of the second. Or again, the diphthong [əʊ] is a realization of the phoneme of *boat* in educated southern British English, but is frequently a realization of the vowel in *boot* in one type of Cockney; however the same number of vowel phonemes occurs in both kinds of English.

Moreover, speakers of different dialects may distribute their phonemes differently in words, as when a speaker from the north of England pronounces *after*, *bath*, and *pass* with /æ/ where a speaker from the south pronounces them with /ɑ:/. Even speakers of the same dialect (as well as those of different dialects) may distribute the same number of phonemes differently among the words they use. In southern British English, some will say *elastic* with /æ/ in the second syllable, others /ɑ:/, and some will say /'ju:nɪzn/ for *unison*, others /'ju:nɪsn/.

Lastly, even individuals are inconsistent; in certain situations, they may change the number of their phonemes, e.g. the occasional use of /m/ in words like *which*; and they may not always use the same phoneme in a particular word or group of words, e.g. the erratic use, in the same person's speech, of /ɒ/ or /ɔ:/ in words like *off*.

We may conclude that a phonemic analysis of a number of varieties of one language is likely to reveal: different coexistent phonemic systems; considerable phonetic discrepancies in the realizations of the phonemes of systems which have an equal number of phonemic categories; variation in the distribution of phonemes in words, even within a community using the same phonemic system; and variation of phoneme distribution, even within the speech of one individual, according to the situation. It is important to remember this likelihood of complication in both the system and its realization, not only in the present situation but also when it is a question of investigating past states of the language. (For a more detailed analysis of variation between dialects, see §7.4.)

5.4 Transcription

The transcription of an utterance (analysed in terms of a linear sequence of sounds) will naturally differ according to whether the aim is to indicate detailed sound values—an ALLOPHONIC (OR NARROW) transcription—or the sequence of significant functional elements—a PHONEMIC (OR BROAD) transcription.

In the former, allophonic, type of transcription, an attempt is made to include a considerable amount of information concerning our knowledge of articulatory

³ Wingate (1982).

activity or our auditory perception of allophonic features. The International Phonetic Alphabet (IPA) provides numerous diacritics for a purpose such as this; e.g. the word *titles* might be transcribed as [tʰä:ëtʰz]. Such a notation would show the affrication and aspiration of the initial [t], the fact that the first element of the diphthong is retracted from Cardinal 4 and is long compared with the second element, which is a retracted Cardinal 2, that the [ʰ] has a back vowel resonance and is partly devoiced in its first stage, and that the final [z] is completely devoiced. Such a notation is relatively explicit and detailed, but gives no more than an impression of the complexity of the utterance as revealed by the various methods of physiological and acoustic investigation. This type of transcription is useful when the focus is on particular details of pronunciation.

In phonemic transcription a different principle operates, that of one symbol per phoneme. Thus a phonemic transcription of the type of English described in this book uses forty-four different symbols. The basis on which an actual symbol is chosen depends on two further principles: (a) use the phonetic symbols of the most frequent allophones, and (b) replace non-roman symbols arising from (a) by roman symbols where these are not already in use. Thus the phonetic symbol for the most common allophone of the phoneme at the beginning of *red* is /ɹ/, but the phonemic transcription replaces /ɹ/ by /r/ on the basis of (b). But in the transcription of vowels romanization (i.e. the principle under (b)) is not completely carried through in this book; e.g. the transcription uses /ɒ/ and /ɔ:/ for the vowels in *cot* and *caught* where it would be possible to use /o/ and /o:/. Transcription of these vowels as used here is called COMPARATIVE PHONEMIC because it allows comparison with vowels in other languages to be made, even though a phonemic transcription is being used. It follows from the principles mentioned above that, even using the IPA, it is possible to construct different sets of symbols for the forty-four symbols of English, although the one used in this book is the most common for the type of English described.

It must be remembered that a phonemic transcription does not by itself indicate how a sequence is to be pronounced. Only if we know the conventions which tell us how a phoneme is to be realized in different positions do we know its correct pronunciation. Nevertheless, a phonemic transcription is particularly useful as a corrective instrument in a language like English, where the orthography does not consistently mirror present-day pronunciation. By now it will have become clear that slant brackets are used for a phonemic transcription, e.g. /tæɪtɪz/, while square brackets indicate an allophonic transcription, e.g. [tʰä:ëtʰz]. Sometimes we may wish to show just the phonetic detail of one segment in an otherwise phonemic transcription. In such cases square brackets must be used, e.g. [tæɪtɪz]. Slant brackets may only be used if the whole sequence is represented phonemically.

5.5 Syllables

The concept of a unit at a higher level than that of the phoneme or sound segment, yet distinct from that of the word or morpheme, has existed since ancient times. It is significant that most alphabets, such as our own, which have as their basis the representation of phonemes by letters (however approximately), have reached this state by way of a form of writing which symbolized a group of sounds—a syllabary.

Indeed, the basis of the writing of many languages, e.g. that of the Semitic group, remains syllabic in this sense. The notion that there exists at this higher level a unit known as the syllable has led to many attempts in recent times to define the term. The best-known approach is that which used to be called prominence theory but is nowadays better known as the sonority hierarchy.

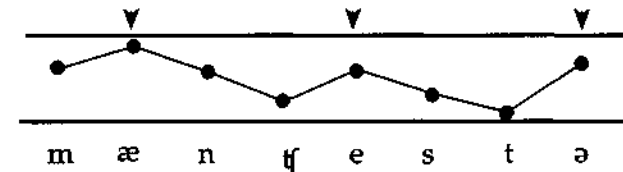
5.5.1 The Sonority Hierarchy

In any utterance some sounds stand as more prominent or sonorous than others, i.e. they are felt by listeners to stand out from their neighbours. Another way of judging the sonority of a sound is to imagine its 'carrying power'. A vowel like [a] clearly has more carrying power than a consonant like [z], which in turn has more carrying power than a [b]. Indeed the last sound, a plosive, has virtually no sonority at all unless followed by a vowel. A sonority scale or hierarchy can be set up which represents the relative sonority of various classes of sound; while there is some argument over some of the details of such a hierarchy, the main elements are not disputed. One version of the hierarchy is as follows (the most sonorous classes are at the top of the scale):

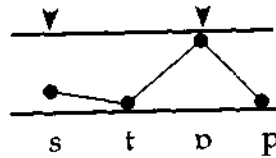
open vowels
close vowels
laterals
nasals
approximants
trills
fricatives
affricates
plosives and flaps

Intermediate vowels are appropriately placed between open and close. Within the last three categories, voiced sounds are more sonorous than voiceless sounds. The class 'approximants' sits somewhat awkwardly in this hierarchy, since the approximants [j,w] are, as already remarked in §4.3.4(5), merely short versions of close vowels in the [i,u] regions. However, the short gliding nature of [j,w] means that they are much less prominent than [i,u] and other approximants like [ɹ,u] have a sonority somewhat less than of laterals and nasals.

Using the sonority hierarchy we can then draw a contour representing the varying prominences of an utterance, e.g.



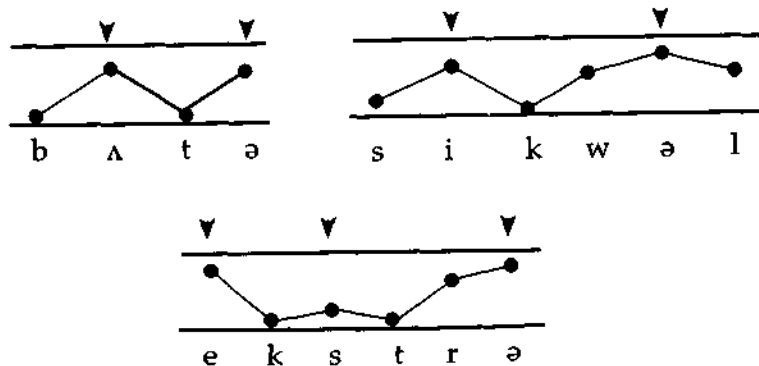
The number of syllables in an utterance equates with the number of peaks of sonority, in this case three (marked with arrows). This accords with native speakers' intuition. However there are some cases where contours plotted with the sonority hierarchy do not produce results which accord with our intuition. Many such cases in English involve /s/ in clusters, as, for example, in *stop*:



The contour of *stop* implies two syllables, while native speaker intuition is certain that there is only one syllable. This suggests that sounds below a certain level on the hierarchy cannot constitute peaks, i.e. that classes from fricatives downwards cannot constitute peaks in English (though the cut-off point may be drawn at different levels in different languages). Formal statements about the clustering possibilities of English consonants sometimes treat /s/ as an appendix which is only a late addition and which may consequently violate restrictions on the sonority of syllables (see §10.9.1(11)).⁴

5.5.2. Syllable Boundaries

In many languages, dividing words and utterances into syllables is a relatively straightforward process (e.g. in Bantu languages, in Japanese, and in French). In other languages, like English, it is not. The sonority hierarchy tells us how many syllables there are in an utterance by showing us a number of peaks of sonority. Such peaks represent the centres of syllables (usually vowels). Conversely, it would seem reasonable for the troughs of sonority to represent the boundaries between syllables. Sounds following the trough would then be in ascending sonority up to the peak and sounds following the peak would be in descending sonority up to the trough. But problems arise, because the hierarchy does not tell us whether to place the trough consonant itself with the preceding or the following syllable; an additional problem is caused by the downgraded, non-peak /s/ mentioned in the previous section. Consider the following examples:



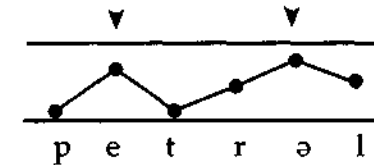
⁴ Giegerich (1992: 147–50).

The question posed by these examples is: to which syllable do we ascribe the intervocalic consonants? Two criteria might reasonably apply in such cases:

- (1) Does a possible syllable division produce sequences of segments which correspond to sequences at the beginnings and ends of words (where we **must** be dealing with syllable beginnings and ends)?
- (2) Does a possible syllable division produce sequences of segments in which the phonetic realization of the segments accords with the way such sequences are realized at the beginnings and ends of words?

If we apply criterion (1) to *butter*, we find that a division into /bΛ/ + /tə/ leaves a syllable end /bΛ/ which never occurs at the ends of words. Therefore /bΛt/ + /ə/ is to be preferred. In the case of *sequel*, both /si:/ + /kwəl/ and /si:k/ + /wəl/ are divisions which accord with the beginnings and ends of words under criterion (1). However, under criterion (2) /si:/ + /kwəl/ accords better with the phonetic realization of segments in those positions: the /w/ in *sequel* is devoiced, and devoicing of /w/ occurs following /p,t,k/ at the beginnings of words and not when the /w/ stands as a single consonant. For *extra* there are three acceptable possibilities by criterion (1): /ek/ + /strə/, /eks/ + /trə/, and /ekst/ + /rə/. However, the /r/ is devoiced in *extra* which suggests that the /r/ must be in the same syllable as the /t/, and hence the third possibility (/ekst/ + /rə/) is ruled out. But there is no non-arbitrary way of deciding between /ek/ + /strə/ and /eks/ + /trə/.

This situation where neither criterion (1) nor criterion (2) helps us to decide between two possibilities of syllable division is not the only type of problem where we have to resort to a non-arbitrary solution. In some cases criterion (1) and criterion (2) actually conflict, e.g. in *petrol*:



/pet/ + /rəl/ accords with the beginnings and ends of words, but does not accord with the devoicing of /r/ present in the word whereas /pe/ + /trəl/ produces the right environment for the devoicing of /r/ (following /t/), yet words do not end in /e/. In such cases we could let the phonetic criterion (2) override (1), or, alternatively, we could regard the /t/ as AMBISYLLABIC, i.e. belonging to both the preceding and the following syllable.

5.6 *Vowel and Consonant*

It was seen in the previous chapter that attempts to arrive at a universal phonetic definition of the two terms, from an articulatory or an auditory standpoint, encounter difficulties as regards certain borderline sounds such as [j,w,r] in English. If, however, the syllable is defined phonologically, i.e. from the point of view of distribution of phonemes, a solution can be given to most of these problems. It will be found that the phonemes of a language usually fall into two classes, those

which are typically central (or nuclear) in the syllable and those which typically occur at the edges (or margins) of syllables. The term 'vowel' can then be applied to those phonemes having the former function and 'consonant' to those having the latter. The English phonemes /j,w,r/, which, according to most phonetic descriptions, are vowel-like, function in the language as consonants, i.e. are marginal in the syllable. The English lateral and nasal sounds are commonly classed phonetically as of the consonantal type because of the complete or partial mouth closure with which they are articulated. From a functional viewpoint, too, they generally behave as consonants, since they are usually marginal in the syllable. Sometimes, however, they operate as a separate peak of sonority, e.g. in *middle* /mɪdl/ and *button* /bʌtn/, and thus function at the centre of syllables. In such occurrences they are referred to as syllabic laterals and nasals.

A further illustration of the consonantal function of /j,w,r/ is provided by the behaviour of the English articles when they combine with words beginning with these phonemes. *The* is pronounced /ði:/ or /ðɪ/ before a vowel and /ðə/ before a consonant; we also have the forms *a* or *an* according to whether a consonant or vowel follows. Since it is normal to pronounce *the yacht, the watch, the rabbit* with /ðə/ and to prefix *a* to *yacht* and *watch* and *rabbit* rather than *an, /j/* and */w/* may be said to behave as if they belong to the consonant class of phonemes, despite their vocalic quality.

It is clear that, if the elements of the utterance are divided into two categories, some units which are assigned to one class according to phonetic criteria may fall into the other class when it is a question of phonological (functional) analysis.

5.7 *Prosodic Features*

We have so far dealt with the description and organization of the qualitative features of utterances. As we have seen in Chapter 3, a sound not only has a quality, whose phonetic nature can be described and whose function in the language can be determined, but also features of length, pitch, and loudness. There may be phonemic oppositions in a language based solely or in part on length differences; alternatively, differences in the length of a phoneme may relate to different contexts, as when English vowels are generally shorter before voiceless consonants than before voiced consonants.

The features of pitch, length, and loudness may contribute to patterns which extend over larger chunks of utterance than the single segment, and when used thus are called suprasegmental, or PROSODIC. Pitch is used to make differences of TONE in tone languages, where a syllable or word consisting of the same segmental sequence has different lexical meanings according to the pitch used with it (e.g. in Chinese). Outside tone languages (and even within tone languages, although to a lesser extent) pitch also makes differences of INTONATION, whereby different pitch contours produce differences of attitudinal or discoursal meaning (discoursal here refers to the way successive chunks of utterances are linked together). While tone is a feature of syllables or words, intonation is a feature of phrases or clauses. Some combination of the features of pitch, length, and loudness will also produce ACCENT, whereby particular syllables are made to stand out from those around them. There are a number of other prosodic features whose linguistic use is far less understood.

These include RHYTHM, the extent to which there is a regular 'beat' in speech; TEMPO (the average conversational tempo of speakers of British English is around four syllables per second); and VOICE QUALITY, which includes both supralaryngeal settings of the mouth and tongue, and laryngeal settings (or phonation types) involving either the vocal cords or the larynx as a whole. Sometimes a voice quality conveys meaning, as when a creaky voice indicates boredom; sometimes a quality is appropriate to a situation, e.g. breathy voice is known as 'bedroom voice' and whispery voice as 'library voice'.⁵

5.8 *Paralinguistic and Extralinguistic Features.*

In addition to prosodic features which spread over more than one segment, there are also PARALINGUISTIC features, which are essentially interruptive rather than co-occurrent. The most common interruptive effect is pause, which functions sometimes as part of the intonation system, where it is one of the indicators of an intonational phrase boundary, but at other times as a hesitation marker. In the latter case a filled pause is often involved, by some combination of [ʔ], [m], and [ə] in southern British, but by other sounds in other dialects and languages (e.g. by an [n] in Russian). Many other paralinguistic effects are more commonly called VOCALIZATIONS: these include single phonemes or sequences of phonemes like [f:] for 'be quiet', [pst] as an attention-getter, and [l] (a reduplicated dental click), for 'irritation' or 'naughty' (often written *tut-tut*) and various conventionalized types of cough and whistle. Since the foreign learner is likely to pause to think of the right word or grammar far more often than the native speaker, it is undoubtedly the hesitation-markers which are the most important feature for him to master. With the acquisition of correct hesitations a foreign learner can dramatically increase his ability to sound like an Englishman.

While prosodic and paralinguistic features are used to convey meaning (although this meaning is in various ways outside the central phonemic system), the term EXTRALINGUISTIC is used for those features over which the speaker has no immediate control. Some of these features may be physical, e.g. sex, age, and larynx size; others may simply be speaker habits, e.g. a particular speaker may always speak with a creaky voice; others may be specific to particular dialects or languages, e.g. speakers of one language may make much more use of an ingressive pulmonic airstream than other languages. Many extralinguistic features are, of course, ones which may also function prosodically or paralinguistically, e.g. breathy voice may be understood prosodically as 'bedroom voice', yet a particular speaker may have this as a constant characteristic of his speech; and voice qualities involving a raised or lowered larynx, while being habitual, may also be interpreted as 'strained' or 'gloomy' respectively.

PART II

The Sounds of English

The Historical Background

6.1 *Phonetic Studies in Britain*

Although linguistic science has made rapid and spectacular progress in the present century, it is not merely in modern times that speech and language have been the object of serious study. Extensive accounts of the pronunciation of Greek and Latin were written two thousand years ago, and in India, at about the same time, there appeared detailed phonological analyses of Sanskrit, which reveal remarkable affinities with modern ways of thought.—‘These early phoneticians speak in fact to the twentieth century rather than to the Middle Ages or even to the mid-nineteenth century . . .’¹In this country, too, printed works containing information of a phonetic kind extend back for at least four hundred years. It is true that the very earliest writers in England rarely had as their main interest a purely phonetic investigation; and the descriptive accounts which they provided are less rigorous and satisfactory, by modern standards, than those of the Indian grammarians. But, by the seventeenth century, we find a considerable body of published work, which is already entirely phonetic in character and which contains observations and theories still adhered to today.

6.1.1 Palsgrave and Salesbury

Some of the first writers whose work we possess were concerned with the relation between the sounds of English and those of another language. Thus John Palsgrave’s French grammar *Lesclaircissement de la Langue Francoyse* (1530) includes a section which deals with the pronunciation of French, much as any modern grammar would. In order to explain the values of the French sounds, Palsgrave compares them with the English. This is done in no objective fashion, and it is not easy for us now to know what precise sound is indicated in either language. But this difficulty of communicating sound values in print—especially those of vowels—was

¹ Allen (1953:7).

one which was shared by all writers until some system of objective evaluation, such as that of the Cardinal Vowels, was devised, and this of course depended on the development of recording techniques. At least John Palsgrave was sufficiently aware of divergent associations of letters and sound to provide some passages of French in a kind of phonetic transcription. Another early writer concerned with pronunciation is William Salesbury, a Welshman, whose *Dictionary in Englyshe and Welshe* (1547) contained comments on the sounds of English. Sound values are indicated by means of a method of transliteration in Welsh or English. Indeed, though grammars of foreign languages published during the next three centuries increasingly attempted more exact description and comparison of sounds, for the great majority of them the section devoted to pronunciation continued to rely mainly on transliteration for indicating approximate values. Even today, grammars of foreign languages frequently make use of this approximate method of 'simulated' pronunciation.

6.1.2 Spelling Reformers: Smith, Hart, Gil

A more important type of phonetic inquiry stemmed from the activities of those who, particularly in the sixteenth and seventeenth centuries, were concerned at the increasing inconsistency of the relationship of Latin letters and the sounds which they represented, especially in English. There had been during the previous five or six centuries great changes of pronunciation, particularly as far as the vowel sounds were concerned, so that letters no longer had their original Latin values. The same sound could be written in a number of ways, or the same spelling do service for several sounds; moreover, the same word might be spelt in different ways by different writers. Thus, four hundred years before the activities of Bernard Shaw and the Simplified Spelling Society, men were aware of the need to bring some order into English spelling. During the four centuries that have elapsed since these early efforts, our pronunciation has continued to evolve without any radical changes of spelling having been made, with the result that today discrepancies between sound and spelling are greater than they have ever been. It can, however, be said that for more than two hundred years our spelling forms, inconsistent though they may be as far as sound symbolization is concerned, have been standardized.

The early spelling reformers were obliged, if they were to propose a more logical relationship of sound and spelling, to investigate the sounds of English. A writer such as Thomas Smith, *De Recta et Emendata Linguae Anglicae Scriptio* (1568), makes many pertinent phonetic comments on such matters as the aspiration of English plosives and the syllabic nature of /n/ and /l/, as well as providing correct descriptions of the articulation of consonants. Yet he, as a phonetician, is overshadowed in the sixteenth century by John Hart, whose most important work, the *Orthographie*, was published in 1569. Besides making out his case for spelling reform and proposing a revised system, Hart describes the organs of speech, defines vowels and consonants (distinguishing between front and back vowels and between voiced and voiceless consonants), and notes the aspiration of voiceless plosives. Of the numerous seventeenth-century orthoepists, only Alexander Gil, *Logonomia Anglica* (1619, 1621), can be compared with Hart on the phonetic level, though even his observations lack the objectivity of Hart's.

6.1.3 Phoneticians: Wallis, Wilkins and Cooper

If the writers mentioned above used phonetic methods of analysis and transcription as a means to their end of devising an improved spelling, there emerged in the seventeenth century a group of writers who were interested in speech and language for their own sake. Because of their preoccupation with detailed analysis of speech activity, the comparative study of the sounds of various languages, the classification of sound types, and the establishment of systematic relationships between the English sounds, they can be said to be the true precursors of modern scientific phoneticians. Two of the most celebrated, John Wallis and Bishop Wilkins, were among the founders of the Royal Society; and, indeed, Isaac Newton, the greatest of the early members of the Society, was interested in phonetic analysis and has left notes of his own linguistic observations. Language was considered a proper object of the attention of the writers of this new scientific age, their view of speech and pronunciation being set against a framework of the universal nature and characteristics of language.

The linguistic fame of John Wallis, primarily a mathematician, spread throughout Europe and lasted into the eighteenth century, his work being copied long after his death. His principal linguistic work, *Grammatica Linguae Anglicanae*, was first published in 1653 and the last authoritative edition appeared in 1699; but other, unauthorized, editions continued to appear in the eighteenth century, the last being dated 1765. Wallis intended his *Grammar* to help foreigners to learn English more easily and also to enable Englishmen to understand more thoroughly the true nature of their language. He admits in his preface that he is not the first to undertake such a task, but claims that he does not seek to fit English into a Latin mould, as most of his predecessors had done, but rather to examine the sounds of English as constituting a system in their own right. By his methods, he says, he has succeeded in teaching not only foreigners to pronounce English correctly but also the deaf and dumb to speak. The introductory part of the work (*Tractatus de Loquela*), besides giving a short history of English, describes in detail the organs of speech and attempts to establish a general system of sound classification which will do service for all languages (illustrations of qualities are taken from French, Welsh, German, Greek, and Hebrew as well as English). Vowels are classified in *Guttural*, *Palatal*, and *Labial* categories, subdivided into *Wide*, *Medium*, and *Narrow* classes. The degrees of aperture are similar to those which are used even today, but the divisions *Guttural*, *Palatal*, and *Labial*, which take into account both the area of raising of the tongue and also lip action, show a confusion of dimensions not to be found in more modern analyses.

Consonants, like vowels, are divided into three classes: *Labial*, *Palatal*, and *Guttural*, being different from vowels in that the airstream from the lungs is obstructed or constricted at some point. Wallis remarks that the airstream may pass entirely through the mouth, almost completely through the nose, or almost equally divided between the mouth and the nose, the position of the uvula determining the difference of direction. Thus nine basic consonantal articulations are postulated. In addition, the airstream may be completely shut off (*Closed* or *Primitive*) or merely constricted (*Open*, *Derivative*, or *Aspirate*), the latter being articulated with a narrow aperture or with a wider, rounder, opening. The 'closed'

consonants (stops) consist of the *mutes* [p,t,k] the *semimutes* [b,d,g], and the *semi-vowels* [m,n,ŋ]. The corresponding 'open' or 'aspirated' consonants are: mute [f,s,x], semi-mutes [v,z,x] or [ɣ]²; and those with a wider opening: mute [ʃ] (again), [θ,h], semi-mutes [w,ð,j]; [l,r] are related to the [d] or [n] articulations; [ʃ,z] are regarded as compound sounds. Wallis's detailed remarks on the pronunciation of English are made in terms of this general system stated in the *Loquela*. It will be seen that such a classification, despite errors and inadequacies which are apparent today, represents a serious attempt at the establishment of universal sound categories. Although the elements of Wallis's system have been quoted briefly here, it should be pointed out that his is merely the most celebrated of a number of similar analyses made at about the same time.

His fellow member of the Royal Society, Bishop John Wilkins, published in 1668 an *Essay towards a Real Character and a Philosophical Language*. Written in English, this work of 454 pages, with a dictionary appended, is of much wider scope than that of Wallis, since it aims at no less than the creation of a universal language, expressed by means of 'marks, which should signifie things, and not words'. Wilkins acknowledges his debt to his contemporary linguists, especially in respect of the account of pronunciation, which forms a comparatively small part of the *Essay*. Wallis, he says, 'seems to me, with greatest accurateness and subtlety to have considered the philosophy of articulate sounds'. Wilkins, too, describes the functions of the speech organs and gives a general classification of the sounds articulated by them; his treatment of consonants is in fact more satisfactory than that of Wallis. He claims that the thirty-four letters which he proposes for his alphabet are sufficient 'to express all those articulate sounds which are commonly known and used in these parts of the world'. In his account of the values of the letters, reference is made not only to European sound systems but also to such little known languages as Armenian, Arabic, Chinese, and Japanese.

Any account of seventeenth-century phoneticians should include the name of Christopher Cooper. Though he did not achieve the great European reputation of Wallis, he is considered by many to be the greatest English phonetician of the century. His work on English pronunciation was first published in 1685 (*Grammatica Linguae Anglicanae*), with an English edition appearing in 1687 (*The English Teacher, or The Discovery of the Art of Teaching and Learning the English Tongue*). A schoolmaster rather than a member of the Royal Society, Cooper was less concerned than many of his contemporaries with the establishment of universal systems. His aim was to describe and give rules for the pronunciation of English for 'Gentlemen, Ladies, Merchants, Tradesmen, Schools and Strangers', rather than to devise a logical system into which the sounds of English and other languages might be fitted. Moreover, he deals with the spelling of English as it exists and does not seek to reform it. The first section of his book is concerned with the description of speech sounds ('The Principles of Speech') and the second part gives rules for the relation of spelling and pronunciation in different contexts. Cooper describes the organs of speech and names those sections of the upper speech tract which are mainly responsible for the articulation of the 'breath': 'guttural, lingual, palatine,

² Wallis's own symbols are here replaced by IPA equivalents, but it is not always clear from Wallis's description which sounds are intended: thus his description of *ch* and *gh* would seem to indicate [x] for both, though in the system one would expect *gh* to mean [ɣ].

dental, labial, linguapalatine, lingua-dental'. Those sounds in the production of which the airstream is 'strained or intercepted' are consonants (classified as semi-vowels, aspirates, semimutes, and mutes), while those in which the airstream is 'freely emitted through the nostrils or the lips' are vowels. He notes that voice, 'made by a tremulous concussion of the larynx', is a characteristic of vowels, semi-vowels, and semi-mutes. His classification of vowels is in terms of *lingual*, *labial*, and *guttural* categories, a somewhat confusing distinction being made between the English long and short vowels. Diphthongs are defined properly as 'a joyning of two vowels in the same syllable, wherein the power of both is kept'. His consonantal classification, with IPA equivalents here, shows: *labial* sounds, subdivided into semivowels [w,v,m], aspirated [ɱ,f,ɱ], semi-mute [b], mute [p]; *lingual* sounds, subdivided into semi-vowels [z,ʒ,ð,n,l,r,j], aspirated [s,f,θ,ʃ,l,t,c], semi-mute [d], mute [t]; *guttural* sounds, subdivided into semi-vowels [ɣ,ŋ,h], aspirated [x,ç,h], semi-mute [g], mute [k]. The second part of the work, dealing with the pronunciation of various English spelling forms, provides more specific information about the pronunciation of English than is to be found in the work of any other writer in this period. Numerous examples are given, e.g. more than three hundred cases of the *-tion* suffix pronounced with [ʃ]; words are listed which have either the same pronunciation with different spellings or the same spellings with different pronunciation; and rules are given for the accentuation of words.

It will be seen from the mention of these few names, chosen from the many who were writing on matters of pronunciation in the seventeenth century and omitting those who were composing spelling books and grammars for foreigners, compiling lists of homophones, and devising systems of shorthand, that there was at this time a surge of scientific and analytical interest in speech and pronunciation such as was not to be repeated until the nineteenth century. It is true that the judgements made were largely intuitive, but this was to remain the case in phonetic research until the second half of the nineteenth century. In their theoretical approach, however, many of these early writers show a preoccupation with classification, systematization, and problems of distribution which is paralleled in the activities of modern linguists.

6.1.4 The Eighteenth Century: Johnson, Sheridan, Walker and Steele

The spirit of general scientific inquiry into speech which characterized a large proportion of the phonetic work of the seventeenth century had, by the eighteenth century, lost much of its original enthusiasm. Prescriptive grammars containing rules for pronunciation continued to be produced in large numbers, and provide us with information concerning the contemporary forms of pronunciation; shorthand systems, too, which show an undiminished popularity, necessitated the analysis of English into its constituent sounds. Yet the main achievement of the century lies in its successful attempt to fix the spelling and pronunciation of the language. Dictionaries had been published in the seventeenth century, but the works having the main stabilizing and standardizing influence on the language were to be the dictionaries of Samuel Johnson (1755), Thomas Sheridan (1780), and John Walker (1791), the last two writers being particularly concerned with the standardization of pronunciation. John Walker, whose dictionary is called by the *Dictionary of*

National Biography 'the statute book of English orthoepy', exerted a great influence on the teaching of English not only in this country but also in America. Moreover, he pays considerable attention in his work to the analysis of intonation, treated perfunctorily by most earlier writers. About the same time Joshua Steele published his *Prosodia Rationalis* (1775–9), in which is presented a system of notation capable of expressing pitch changes, stress, and rate of delivery. (Steele is celebrated for his detailed analysis of a soliloquy delivered by David Garrick.)

We have, in fact, been dealing up to now with two types of work on pronunciation, which, especially in the eighteenth century, came to be confused: on the one hand, and in the minority, the books which laid emphasis on description, analysis, and classification; on the other, the books which were mainly normative and continued the tradition of 'rhetoric'. That part of rhetoric known as 'elocution' originally referred to the style and form of speech, 'the garnishing of speech', but in the eighteenth century the term was increasingly applied to the method of delivery. It was not until the nineteenth century that a clear distinction was made between the aesthetic judgements upon which elocution largely relies and the objective descriptive statements which form the basis of phonetic analysis. Until such a distinction was explicitly made, advances in phonetic techniques have to be disentangled from a mass of irrelevant opinion.

6.1.5 The Nineteenth Century: Pitman, Ellis, Bell, Sweet

In the nineteenth century the traditional British preoccupation with phonetic notation and the simplification of English spelling continued. Isaac Pitman (1813–97), whose system of shorthand is so widely used today, and Alexander J. Ellis (1814–90), concerned at the difficulties which our spelling presented to English children as well as to foreigners, devised an alphabet, Phonotype,³ which conformed to a phonetic analysis of English and yet remained based upon the Latin characters. They were supported by the Phonographic Society and published a journal which eventually (1848) was named the *Phonetic Journal*. Ellis, however, developed other types of alphabet, notably Glossic, which is essentially an adaptation of traditional spelling, and Palaeotype, which used conventional letter shapes but in a great variety of type, so that fine shades of sound could be symbolized. This latter alphabet was put to good use by Ellis in his historical and dialectal studies; but not only is the precise value to be attached to a letter not always easily determined—because of the method of reference to sounds in languages—but also the complexity of the system renders it difficult for the reader to assimilate. Ellis's work on notation, however, largely inspired the 'Broad' and 'Narrow' romic transcriptions of the great Henry Sweet (1845–1912). In 1867 Alexander Melville Bell, father of Alexander Graham Bell (the inventor of the telephone), published his book *Visible Speech*, while a lecturer on speech in the University of London. This remarkable work set out to classify all the sounds capable of being articulated by the human speech organs and to allot a systematic and related series of symbols to the sounds. The unfamiliarity of the invented symbol shapes was no doubt responsible for the fact that this means of notation has never been widely used in purely phonetic

work, but its value was for many years demonstrated, especially in America, as a system applied to the teaching of the deaf.

Although, in referring to these writers, emphasis has been laid on their contribution to the development of phonetic transcription, their published work covers every aspect of speech activity. Bell's interests, in his forty-nine publications, lay mainly in the field of elocution and the description of articulatory processes. But Ellis and Sweet applied the techniques of phonetic analysis both to the description of contemporary pronunciation and also to the whole field of historical phonological investigation. Ellis, in fact, will be chiefly remembered for his massive work *On Early English Pronunciation*, published in five volumes between 1869 and 1889. In these volumes Ellis traces the history of English pronunciation and, at the same time, contributes descriptive phonetic studies of contemporary dialects. It is not surprising that a work of such enormous scope should since have been found to be inadequate in many respects, but it cannot be denied that Ellis was a great pioneer in the application of objective techniques to the description of past and present states of the language. Although his assessment of the value of many grammarians from the sixteenth century onwards was often faulty, he initiated a study of their work which has continued unabated to this day. Henry Sweet, a greater phonologist and scholar, applied stringent phonetic techniques to all his work, so that, whether it be a question of phonetic theory or the history of English or the description of a language such as Welsh or Danish, his basic approach and the majority of his conclusions remain valid today. He belongs as much to the twentieth century as to the nineteenth, and his influence is clearly to be seen in the work of Daniel Jones, who dominated British phonetics in the first half of this century.

This brief and selective outline goes some way towards revealing a line of phonetic inquiry which has been continuous in England from the sixteenth century to the present day. The techniques for describing speech and language have become progressively more objective, modern instrumental methods for physiological and physical investigation (now supported by digital computers) providing the latest stage in the process. A problem confronting linguists of today concerns the correlation of concrete data, which is being accumulated in great detail in computer databases, with abstract linguistic realities, many of which have for centuries been implicit in the work of writers on language.

6.2 Sound Change

The language spoken in England has undergone very striking changes during the last thousand years, changes which have affected every aspect of the language, its morphology, syntax, and vocabulary as well as its pronunciation. Old English⁴ is so different from present-day English from every point of view that it is unintelligible to the modern Englishman either in its written form or in a reconstructed spoken form; Chaucer's poetry presents difficulties in print and, when read in what is presumed to be the pronunciation of the fourteenth century, offers a sound pattern

3 See Kelly (1981).

4 The following abbreviations will henceforward be used: OE: Old English (up to c. AD 1100); ME: Middle English (c. 1100–1450); eModE: early Modern English (c. 1450–1600); PresE: present-day English; AN: AngloNorman; OF: Old French.

which it is not easy for the modern listener to interpret; even Shakespeare, though phonetically not far removed from ourselves, raises problems of syntax and meaning.

The pronunciation of a language seems to be subject to a continuous and inevitable process of change. Indeed, it would be surprising if a means of communication, handed on orally from one generation to another, showed no variation over the centuries. It is not difficult to find examples of changes which are taking place in our own times—e.g. the final vowel in words like *city* has become closer and tenser, and the vowel in *sad* has become more open amongst young people in the south of England today than they are in the pronunciation of older people. A change of a different kind—the use of another phoneme in a class of words—is illustrated by the case of words such as *poor* and *sure*; these tend to be said by the older generation with /uə/, whereas the younger generation much more commonly uses /ɔ:/. At any given moment, therefore, we must expect several pronunciations to be current, representing at least the older, traditional, forms and the new tendencies.

Today there are a number of reasons why we might expect these processes of change to operate less rapidly. The fact that communication throughout the whole country is easy, the spread of universal education and the resultant consciousness of the printed word, the constant impact of broadcasting with its tacit imposition of a standard speech, these are all influences which are likely to apply a brake to change in pronunciation. They are, however, factors which have operated only in comparatively recent times. In former stages of the development of English, there was no mass, nationwide, influence likely to lead to stability and levelling. Printing, it is true, has been with us for four hundred years, but the wide dissemination of books, as of education, is a modern development. Indeed, as we have seen, the spelling of English, even in printed books, was not finally standardized until the eighteenth century. With such freedom from restraint, especially before the eighteenth century, it was not unexpected that there were considerable changes of pronunciation.

6.2.1 Types of Change

(1) The most important kind of change tends to affect a phoneme in all its occurrences. Such changes, not usually being set in motion by any immediate, outside, influence, are in this sense independent; they are called **INTERNAL ISOLATIVE** changes. Thus the ME realization of the phoneme in a word such as *house* had the sound [u:], which has generally become [aʊ] in modern English; similarly, the ME vowel phoneme having a value of the [a:] type, as in a word such as *name*, is in most cases realized as a kind of [eɪ] in PresE. Changes of this type apply particularly to the English vowel system, which underwent a remarkable evolution of values, known as the **GREAT VOWEL SHIFT**, during the centuries preceding the modern period.

(2) Another kind of change is that which is brought about by the occurrence of phonemes in particular contexts—a dependent change, called **INTERNAL COMBINATIVE**. Thus, the phoneme in *mice*, having now the sound [aɪ], results from an earlier [i:] by means of an isolative change; but this [i:] sound in [mi:s] arose as a result of a combinative process of **VOWEL HARMONY**, or **i-mutation**, through the

stages [mu:si], [my:si] > [mi:s], where the change [u:] > [y:] can be explained by the fronting of [u:] under the influence of the [i:] of the following syllable. Such a combinative change belongs to OE, but a more recent change of this type is exemplified by words such as *swan*. This word was probably pronounced [swan] or [swæn] in about 1600, but the [w] sound has rounded and retracted the vowel to give the modern form [swɒn]. The large majority of earlier [w] + [a] sequences have now given [w] + [ɒ], or [ɔ:], by reason of this combinative change affecting this particular sound sequence, e.g. *want*, *quality*, *war*, *water*.

(3) Some changes are neither independent nor dependent upon the phonetic context; they may be said to be **EXTERNAL** to the main line of evolution. Thus it was fashionable in Elizabethan times to pronounce such words as *servant* and *heard* with [æɪ] or [ar], perhaps originally a dialect form, rather than with [ɛr], the regular form of development; these words, with some exceptions such as *clerk*, have reverted to the normal development of ME [ɛr] > [ɜ:] rather than [ɑ:]. It was also fashionable to pronounce the termination *-ing* as [ɪŋ], only now retained as a special form of affectation or in some dialects. Such changes, involving a change of distribution of phonemes among words and morphemes, do not affect the phonemic system of the language. The introduction of foreign words may, however, at least temporarily and in the speech of a restricted number of individuals, disturb the number of phonemes or their distribution as regards position in the word. Thus, if the French word *beige* is used in English with the pronunciation /beɪʒ/, we have a case of a final /ʒ/ previously unknown in English words; or again, if *restaurant* is pronounced with any kind of nasalized vowel in the last syllable, the possibility of a new kind of vocalic opposition is introduced into the language. However, such foreign borrowings generally tend to conform to the English system: words with a final French /ʒ/, such as *prestige* or *camouflage* may be realized in the English form with /dʒ/, and a word with a nasalized vowel like *restaurant* will be normalized to /ˈrestɔːrɒŋ/, /ˈrestɔːrɒnt/, or /ˈrestɔːrɒnt/.

(4) In addition to changes of quality, there have also to be taken into account changes involving **LENGTH** and **ACCENTUAL PATTERN**. Thus the vowel in such words as *path*, *half*, *pass*, still short three hundred years ago, is now long in the south of England. Or again, the vowels in *good*, *book* and *breath*, *death*, once long, are now relatively short. Changes of accent are particularly striking in the case of words which have come into the language from French: in ME, such words as *village* or *necessary* retained their accent on the penultimate syllable—/vɪˈla:ʒə/ and /nesəˈsɑ:rɪ/. Now, the accent has shifted to an earlier syllable, together with associated changes of quality—/ˈvɪlɪdʒ/, /ˈnesəsəri/ (the latter may retain the ME pattern in American English). Later borrowings, or those in less common use, often retain the French accentual pattern—thus *hotel* or *machine*, have the accent on the final syllable, whereas, if they had conformed to the English system, we might have had such modern forms as /ˈhəʊtəl/ and /ˈmæʃɪn/ or /ˈmeɪʃɪn/, in the same way that the thoroughly anglicized form of *garage* gives /ˈgærɪdʒ/. (See §7.5 on current changes.)

6.2.2 Rate and Route of Vowel Change

The English vowels have been subject to more striking changes than have the consonants. This is not surprising, for a consonantal articulation usually involves

an approximation of organs which can be felt; such an articulation tends to be more stable, in that it is more easily identified and transmitted more exactly from one generation to another. Changes in the consonantal system comparatively rarely involve a modification of sound (an example of such a modification would be the affrication, for combinative reasons, of the OE palatal plosives [c,ɟ] to [tʃ,ɔʒ] as in *church* < OE *cirice* and *bridge* < OE *brycg*). Far more common is the type of distributional change involving the conferment of phonemic status on an existing sound (e.g. [v,ð,z], allophones of /f,θ,s/ in OE, later obtain contrastive, phonemic, significance), the disappearance of an allophone (e.g. postvocalic [x] and [ç] in such words as *brought* and *right* were largely lost in the south of England by the seventeenth century), or the insertion of an existing phoneme in a particular class of words (e.g. the initial /l/ in words of French origin such as *herb*, *homage*). Whether it is a question of consonantal change, loss, or addition, it is usually possible to explain the type of modification which has taken place and the approximate period during which it occurred.

A modification of vowel quality will, however, result from very slight changes of tongue or lip position, and there may be a series of imperceptible gradations before an appreciable quality change is evident (or is capable of being expressed by means of the Latin vowel letters). It is particularly difficult to assess rate and phonetic route of change in the case of those internal independent vowel changes which affect a phoneme throughout the language. It is known, for instance, that the modern homophones *meet* and *meat* had in ME different vowel forms, approximately of the value [e:] and [ɛ:]. The [e:] vowel of *meet* became [i:] by about 1500, and it might be postulated that by a process of gradual change the [ɛ:] of *meat* first closed to [e:] and then, by the eighteenth century, coalesced with the [i:] in *meet*. The available evidence, however, suggests that the change [ɛ:] > [i:] may not have been either simple or gradual, but that two pronunciations existed side by side for a long period (the conservative [ɛ:] beside another form [i:] which had resulted from an early coalescence with the *meet* vowel). In other vowel changes, it may be agreed that the change was gradual, but it is difficult to date precisely the stages of development. Thus the modern /aɪ/ of *time* results from a ME [i:] value; it is clear that the change has been one of progressive, widening diphthongization, but there may have been a period of incipient diphthongization when there was hesitation between the pure vowel [i:] and some such diphthong as [ii] or [əi]. It is well to remember, therefore, that at any particular time in history there are likely to be a number of different, coexistent realizations of vowel phonemes, not only between regions but also between generations and social groups. An example of such variety in modern English is provided by the vowel at the end of *city*, which in the south of England may be rendered as [ɪ] by the older generation and as something more like [i] by younger people. The speech of any community may, therefore, be said to reflect the pronunciation of the previous century and to anticipate that of the next.

6.2.3 Sound Change and the Linguistic System

It is convenient to study sound change in terms of the development of particular phonemes or sounds, but it is misleading to ignore the relationship of the sound units to the system within which they function and which may, in fact, not be

changing. In other words, although there may be considerable qualitative changes, the number and pattern of the terms within the system may show relative stability. The ME /i:/ phoneme, for instance, is now realized as [aɪ], but there is still a phonemic opposition which contrasts such words as *time*, *team*, *tame*, *term*, *tomb*, and, in any case, a new phoneme /i:/ has emerged in words of the *team* type. On the other hand, the system may change because a sound, without itself changing, may receive a new, phonemic, value; e.g. the sound [ŋ] has always existed in English as a realization of /n/ followed by the velars /k/ or /g/, but when the final /g/ in a word like *sing* was no longer pronounced, /ŋ/ contrasted significantly with /n/ and /m/, e.g. *ram*, *ran*, and *rang*.

Since the system of our language consists of a framework of significant oppositions by means of which we communicate, it may be assumed that there is a tendency for the system to remain stable, the loss of an opposition involving a possibility of confusion. In fact, of course, the redundancy of English is such that some degree of neutralization of phonemes is easily tolerated: today, few speakers in the south of England distinguish *saw* and *sore* by means of an opposition /ɔ:/- /ɔə/, yet the loss of the /ɔə/ diphthong is no impediment to communication. An example of an earlier coalescence of vowel phonemes is that illustrated by the homophony of *meet* and *meat*. On the other hand, new oppositions may emerge in the language, e.g. the phonemes /v,ð,z,ŋ/, as we have seen. Nevertheless, despite the adjustments in the number of phonemes which have taken place, the history of the English sound system displays, over the last 1,000 years, a considerable degree of stability.

Though the relationships within the system may tend to remain stable, a change of phonetic realization of any phoneme is likely to have qualitative repercussions throughout the system. Such a disturbance may be observed in modern English. The phonetic relationship of the vowel phonemes in *set* and *sat*, in one type of pronunciation, is of a front vowel between close-mid and open-mid to a front vowel between open-mid and open. If, however, the vowel of *sat* has a closer articulation than that described, that of *set* must be raised too. A limit of raising is imposed by the presence of *sit* and *seat*, for it is not possible to raise the vowel of *sit* to any extent without danger of confusion with that of *seat*, unless the latter vowel becomes strongly diphthongal. (It may be objected that a quantitative as well as qualitative difference distinguishes /i:/ from /ɪ/; but in the examples given—*seat* and *sit*—the phonetic context imposes a quantity on /i:/ which is practically the same as that of /ɪ/. If /ɪ/ were too close to the region of /i:/, the opposition would be maintained only by realizing /i:/ as fully long at the expense of the shortening influence of the final /t/ (or by a process such as diphthongization.) Alternatively, if the vowel phoneme of *sat* is realized as a front open vowel, as in many English regional dialects, the vocalic area in which the phoneme of *set* can be realized becomes more extensive; in fact, in those kinds of English where this occurs, the vowel in *set* tends to be open-mid variety. Such considerations of the phonetic relationship of phonemes have a relevance in the historical, DIACHRONIC study of English. In ME there were, for instance, four long vowels in the front region—/i:,e:,ɛ:,a:/. By 1600 /i:/ had diphthongized and the remaining vowels closed up. Such a movement may have been caused by pressure upwards from /a:/ or by the creation of an empty space brought about by the diphthongization of the pure vowel /i:/.

Although, therefore, it is often convenient in diachronic studies to investigate the development of individual phonemes in terms of the quality of their realization, it is clear that many sound changes can be explained only by reference to a readjustment of the phonetic relationships of the phonemes of the system as a whole. Moreover, any particular point in the development of the sound system of a language is not simply to be considered as a stage in the process of change of a number of sound units but rather as the presentation of the functioning of a system at a certain historical moment. The primary significance of the sounds of modern English is their function in the system of today; in the same way, the English sounds of 1600 are to be viewed in terms not only of their past and future forms but also of their contemporary, SYNCHRONIC relationships and functions.

Some sound changes are, indeed, the result of an influence which applies to the system as a whole. Those drastic changes of vowel quality known as the Great Vowel Shift mainly affect vowels in accented syllables. But vowels in most unaccented syllables (especially those in word-final positions) have undergone, in the last thousand years, an equally striking, though different, type of change. Henry Sweet has called OE the period of full endings, *stanas* being realized as [ˈstɑ:nɑs]; ME, the period of levelled endings, when *stones* was pronounced [ˈstɔ:nəs]; and eModE and later English, the period of lost endings, when *stones* is [sto:nz], [stəʊnz]. There is, therefore, a general tendency for all unaccented vowels to shorten (if long) and to gravitate towards the weak centralized vowels [ɪ] or [ə], or sometimes [ʊ], if not to disappear altogether. This fact accounts for the high frequency of occurrence of [ɪ] and [ə] in PresE and for the complete elision of many vowels in unaccented syllables in rapid colloquial speech, e.g. *suppose* [spəʊz], *probably* [p.rɒb.əb.li].

6.2.4 Sources of Evidence for Reconstruction

Whether our aim is to reconstruct the phonological system of English at any particular moment in history or to estimate the nature of the development affecting particular phonemes, it is necessary to establish the sound values which were used in the pronunciation of the language—relative values in the case of the system, absolute values as far as possible in the case of sound development. An investigation of the phonological structure of PresE would have to include direct observation of its phonetic features. For this purpose, future generations will have the benefit of recordings of the speech of today. Obviously, this type of evidence cannot be used for the reconstruction of past states of the spoken language. The further back we go into history the scantier the evidence of spoken forms becomes. Our conclusions will, therefore, be based on information mostly of an indirect kind; yet such is the agreement generally amongst the various types of evidence that the broad lines of sound change can be conjectured with reasonable certainty.

(1) *Theoretical paths of development.* If, in dealing with the changing realization of a particular phoneme, we can be reasonably sure of its sound value at two points in history, we can, from our knowledge of phonetic possibilities and probabilities, infer theoretically the intervening stages of development. We can, of course, be sure of the pronunciation of PresE. If, then, the evidence suggested unequivocally that, for instance, the vowel in *home* was pronounced as [ɑ:] in OE, the development to

be described and accounted for would be [ɑ:] > [əʊ]. It is likely that the articulation has always involved the back, rather than the front, of the tongue; the change has clearly meant a closing of the tongue position, to which at some stage there has been added a gliding (diphthongal) movement. We might, therefore, postulate such developments as [ɑ:>ɑʊ>əʊ] or [ɔ:>ɔʊ>əʊ]. The available evidence will then confirm or refute the hypothesis—in this case the second solution being more in keeping with the information. Such recognition of phonetic probabilities will always be implicit in the tracing of change. It must be considered unlikely that [ɑ:] on its way to [əʊ] or [əʊ] would have passed through a stage of front articulation, without any combinative influence. Nevertheless, the possibility of a type of change which is not the most probable theoretically must never be excluded. The rounded close-mid back ME [ɣ] developed by the nineteenth century to an unrounded open-mid centralized back [ɤ]; and in the London area this vowel has now become more open and more front [ɛ]. Yet, at the same time, there is a tendency to make the vowel in *sad* more open. There is here a potential conflict, and the future development of these vowels is uncertain. It would, therefore, be dangerous to predict, merely according to phonetic probabilities, the way our present sound system will develop.

(2) *Old English.* It is most important in an investigation of the development of English sounds over the last thousand years that the pronunciation of OE should be established with some certainty. If this can be done, we shall have a 'starting-point' for the phonetic route of change to PresE. The term Old English, however, spans a period of some four hundred years from about AD 700 AD 1100. Moreover, the invasion of the Angles, Saxons, and Jutes in the fifth and sixth centuries introduced four separate varieties of English: the Angles, in the Midlands, north-east England, and the south of Scotland, using types of English known as MERCIAN and NORTH-UMBRIAN (or, in general terms, ANGLIAN); the Saxons, in the south and south-west, using the WEST SAXON dialect; and the Jutes, settling mainly in the region of Kent and using a dialect called KENTISH. Of the four dialects, West Saxon, which was to become a kind of standard language, is the one about which most is known from the extant texts. In its later form—that in use between about AD 900 and AD 1100—it is referred to as Classical OE.

The broad lines of the pronunciation of this language can be conjectured from a comparison of the development of the other members of the West Germanic group of languages to which it is related. But by far the most explicit evidence concerning its sounds is to be inferred from the alphabet in which it is written. The earlier runic spelling was replaced by a form of the Latin alphabet. This alphabet was probably introduced into the country in the seventh century by Irish missionaries. It can be assumed, therefore, that the sounds of OE were represented as far as possible by the Latin letters with their Latin values, with some modifications of an Irish kind. A great deal is known about the pronunciation of Vulgar Latin, whose sound system had much in common with that of modern Italian. If an Italian, knowing no English, were today asked to write down with his own spelling the PresE pronunciation of the word *milk* [mɪk], he would have no difficulty in representing the first sound, which he could spell as *m*; the vowel [ɪ] might, however, seem to him to resemble the sound he would write in Italian as *e* rather than as *i*; the 'dark' [ɨ] would appear to have a back vowel glide accompanying it, requiring a spelling such as *ol*; and, since he has no *k* letter, he would spell the final [k] as *c*. His transcription

of the word might, therefore, be *meolc*, which is, in fact, a West Saxon spelling of the word now written *milk*. This is a fortuitous example, and must not be taken to suggest that OE was pronounced in the same way as PresE. But it does demonstrate that OE spellings, which may appear to be very different, are often less surprising when we keep in mind the Latin values originally attached to the letters.

Sometimes the simple forms of the Latin alphabet were evidently inadequate for representing the English sound: thus the joined form *æ* was used to symbolize a sound between C[a] and C[ɛ]; the sounds [θ] and [ð] were written in the earlier manuscripts as *th* initially and *d* medially and finally in a word, and later as [ð] or the rune þ, regardless of the sound's position in the word or its voiced or voiceless quality; the rune þ frequently replaced the earlier *u* or *uu*. The vowel values of the OE system were particularly difficult to represent with the five Latin vowel letters. Sometimes the spelling used hesitated between two letters: thus the vowel of *mann*, probably of a C[a] or [ɔ] quality, was written either with *a* or *o*, indicating a vowel between the unrounded open central value of the Latin letter *a* and the rounded open-mid to close-mid back value of *o*. Unaccented vowels, too, already beginning to be obscured and levelled, presented a problem to the scribes, the Latin alphabet offering no way of showing a central vowel of the [ə] type. Unaccented *æ*, *e*, and *i* soon began all to be written as *e*, and unaccented *a*, *u*, *o* later tended to be used indifferently, indicating that the vowel distinction was being lost. A diphthong such as the one written as *ea* must probably be interpreted as a glide to a central [ə] quality.

Quantity is often shown in the case of vowels by doubling the letter or by the use of an accent and in the case of consonants by doubling the letter. The accent in a word is also sometimes shown by the use of a mark; but, in any case, it is agreed, from a comparison of the West Germanic languages, that the word accent in OE fell generally on the first syllable of words, with the exception of certain compounds.

The written form of OE provides us, therefore, with considerable information concerning the language's pronunciation; we have a working hypothesis from which to begin our investigations. The study of later forms of English will often, in fact, confirm that the OE pronunciation postulated from the spelling and the comparison of Germanic languages is the only one from which later forms can be expected to have developed.

(3) *Middle English*. Spelling forms can also help us to deduce the pronunciation of the ME period, roughly AD 1100–1450. Generally speaking, it may be said that the letters still had their Latin values and that those letters which were written were meant to be sounded. Thus, the initial *k* in a word such as *knokke* was still pronounced and the vowel in *time* would have an [i] quality. This persistence of Latin values in spelling was no doubt due to the influence of the Church, which was still the centre of teaching and writing, and the absence of a thoroughly standardized spelling accounts for its predominantly phonetic character. However, English spelling was modified by French influences. Notably, the French *ch* spelling was introduced to represent the [ʃ] sound in a word such as *chin* (formerly spelt *cinn*), where the new spelling form indicates no change of pronunciation; in addition *ou*, or *ow*, represents the sound [u], formerly written *u*, e.g. *hous*, in OE *hus*. The simple *u* spelling was retained to express both the French sound [y] in words like *duke* and *fortune* and the OE short [u] sound, though this latter sound is often written as *o*,

especially when juxtaposed with letters of the *w*, *m*, *n* type, e.g. *wonne* rather than *wunne*, to avoid confusion between the letter shapes.

Rhymes, too, have their value, especially as, in this period, they are likely to have been satisfactory to the ear as well as to the eye—in the whole of Chaucer's work, for instance, there are very few rhymes which appear to involve the pairing of different vowel sounds. Nevertheless, evidence from rhymes is valueless unless it is possible to be certain, from other sources of evidence, of the pronunciation of one member of the pair. Thus, the Chaucerian rhyme *par cas* :: *was*, because we can be sure that the French word *cas* had a vowel of the [a] quality, is evidence to confirm the view that the [w] of *was* had not yet retracted and rounded the vowel to [ɔ] and, the final *s* in the two words being still likely to represent [s], that the word was probably pronounced [was].

Again, words imported from French can give us information concerning the timing of sound changes. Thus French words such as *age* and *couch*, which we know from French sources had [a:] and [u:] at the time of their introduction into English, fell in with the English vowel development [a:] > [ɛr] and [u:] > [au] in words like *name* and *house*; we can conclude, therefore, that at the time the French words came into the language the [a:] and [u:] vowels had not begun their change.

Moreover, after the ME period, as we shall see, a great deal of direct evidence is available to us, so that our conjectures from about 1500 onwards can be made with considerable certainty. We may often, therefore, be able to deduce from our knowledge of pronunciation in the sixteenth century, the stage probably reached in the ME period in the development of a sound from OE. The OE [i:] sound in *time*, for example, was beginning to be diphthongized generally very early in the sixteenth century. It is reasonable to suppose (even if other evidence to support the theory did not exist) that *time* still had a relatively pure [i:] for much of the ME period.

Finally, the metre of verse reveals the accent of words. It is for this reason that we know that French words, in Chaucer's verse, generally retained their original accentual pattern, e.g. *courage* [ku'ra:ʒə], and that the accent shift in these cases is a phenomenon of at least late ME.

(4) *Early Modern English*. The same sources of evidence which we have already considered remain available for the eModE period, roughly AD 1450–1600. The introduction of printing brought standardization of spelling, and already the spoken and written forms of the language were beginning to diverge. But individuals, especially in their private correspondence, often used spellings of a largely phonetic kind, in the same unsophisticated and logical way that children still do. If a modern child writes *He must have gone* as *He must of gone*, he is only representing the phonetic identity of the weak forms of *have* and *of* ([əv]), an identity which he will learn to ignore when he adopts the conventional spelling distinction. In the same way, if fifteenth- and sixteenth-century spellings show the word *sweet* occasionally written as *swit*, it may be assumed that this original ME [e:] was by now so close that it could be represented by *i* with its Latin value. Or again, the spelling form *sarvant* instead of *servant* reflects an open type of vowel in the first syllable which was current throughout the eModE period in such words. Moreover, the conventional adoption of an unphonetic spelling can sometimes provide us with positive evidence as to its value: thus, when words like *delight* (formerly *delite*) began to be spelt with *gh*, this spelling form *gh* clearly no longer had the

consonantal fricative value which it had formerly represented in *light*, since there never was a consonantal sound between the vowel and final [t] in *delight*. We may conclude, therefore, that *gh* no longer had its former phonetic significance in words such as *light*. Care must, of course, be taken to identify the increasing number of learned or technical spellings adopted by printers. The initial letter group *gh* in *ghost* (OE *gast*) indicates no change in pronunciation—*goose* was also sometimes spelt *ghoose* in this period. Again, spellings which aim at revealing the etymology (true or false) of a word must usually be discarded as phonetically valueless, e.g. *debt*, *island*. Thus from the writings of individuals some general indications concerning sound changes may be gathered and used to supplement evidence derived from other sources.

Rhymes, too, continue to be useful as complementary evidence. A rhyme such as *night* :: *white* confirms the view that post-vocalic *gh* no longer had a consonantal value; or again, *can* :: *swan* suggests that the rounding of [a] after [w] had not yet taken place. Yet, just as in the case of ME, rhymes must be treated with caution, more particularly as eye-rhymes were doubtless beginning to become more prevalent. In Elizabethan literature, however, additional evidence is afforded by the frequent use of puns, which usually rely for their effect upon similarities, if not identities, of phonetic value. Shakespeare, for instance, plays on the phonetic identity of such pairs as *suitor*, *shooter* (both capable of being pronounced [ʃu:tər]) and *known*, *none* (both [no:n]); such puns suggest that the pronunciation of the two words was commonly sufficiently close to make an immediate impression upon an audience.

The most important and fruitful evidence for this period is, however, of a direct kind. It is provided by the published works of the contemporary grammarians, orthoepists, and schoolmasters, some of whom have been mentioned in §6.1. They are of unequal value and their statements have often to be interpreted in the light of other evidence; yet they provide us with the first direct descriptive accounts of the pronunciation of English. From the sixteenth century onwards, our conclusions rely more and more on their descriptive statements and less on clues of an indirect kind. Sometimes there appears to be a conflict between the phonetic probabilities, the statements of grammarians, and evidence from other sources. Frequently the solution must be that there existed at any time a variety of current pronunciations, resulting from differences of dialect, generation, fashion, and place in society, in the same way that a description of PresE (even that of a restricted area such as the south of England) would have to take into account a large number of variants.

The following representative systems are conjectures of one possible set of phonemes current in the periods in question.

6.2.5 The Classical Old English Sound System

<i>Vowels</i>	i:,i:y:ɣ u:,u e:,e o:,ɔ æ:, æ ɑ:,ɑ (allophone [ɒ] before nasal consonants) [ə] occurs in certain weakly accented syllables
<i>Diphthongs</i>	e:ə, eə; e:ə, eə
<i>Consonants</i>	p, b, t, d, k, g (allophone [ɣ])

ʃ, ʒ
m, n (allophone [ŋ] before velar consonants)
l, r
f, θ, s (medial allophones [v, ð, z])
ʃ, h (allophones [x, ç])
j, w

Consonants may be long or short.

The spellings *hn*, *hl*, *hr*, *hw* may be interpreted as phoneme sequences /h/ + [n, l, r, w]; alternatively, if it is assumed that *h* is here an indication of voiceless [h̥, l̥, r̥, w̥], these four sounds may be counted as contrastive, i.e. of phonemic status.

Text (St John, Chapter 14, verses 22, 23)

- 22 ju:das kwæθ to: him. næs nɑ: se: skariot driçtən, hwæt is jəwərdən
θæt θu: wilt θe: sylfnə jəswo:teljən ʊs næs middane:rdə.
23 se: hæ:lænd ʊndswarɔdə ʊnd kwæθ him; jif hwa: me: lʊvəθ he: hult
mi:nə spræ:ʃə ʊnd mi:n fædər lʊvəθ hi:nə ʊnd we: kʊmæθ to: him
ʊnd we: wyrkiəθ e:ardungsto:wə mid him.

Authorized Version

- 22 Judas saith unto him, not Iscarioth, Lord, how is it that thou wilt
manifest thyself unto us, and not unto the world?
23 Jesus answered and said unto him, If a man love me, he will keep my
words; and my Father will love him, and we will come to him, and
make our abode with him.

6.2.6 The Middle English Sound System

<i>Vowels</i>	i:,i u:,u e: o: e:,e ɔ:,ɔ ɑ:,ɑ ɑ: [ə] occurs in unaccented syllables
<i>Diphthongs</i>	eɪ,(æi),ɔi, iu,(eu), eu,ɔu,(au)
<i>Consonants</i>	p, b, t, d, k, g, ʃ, ʒ m, n (allophone [ŋ] before velar consonants) l, r f, v, θ, ð, s, z, ʃ, h (allophones [x, ç]) j, w (allophone [w] after /h/)

Text (from the *Prologue to the Canterbury Tales*)⁵

hwan θat a:pril, wiθ his ʃu:rəs so:tə
θə dru:xt ɔf martʃ hæθ pɛrsəd to: ðə ro:tə,
and bɑ:ðəd e:vri væin in swiʃ liku:r
ɔf hwɪʃ vertiu endʒendərd is θə flu:r.

⁵ The type of transcription given here is slightly archaic for Chaucer's pronunciation; e.g. long consonants were probably lost in later ME and such words as *and*, *that* would have had a weak vowel.

hwan zefirus e:k wiθ his swe:tə bræ:θ
 inspi:rəd hæθ in e:vri holt and he:θ
 θə tændər krɒppəs, and ðə jʊŋgə sʊnnə
 hæθ in ðə rəm his halvə kɔrs ɪrʊnnə,
 and smɑ:lə fu:ləs mɑ:kən melɒdi:ə
 θæt sle:pən ɑ:l ðə niçt wiθ ɔ:pən i:ə—
 sɔ: pri:kəθ hem nɑ:tiur in hɪr kura:dʒəs—
 θæn l:ŋgən fɔlk tɔ: gɔ:n ɔn pilgrɪmɑ:dʒəs.

6.2.7 The Early Modern English Sound System

<i>Vowels</i>	i:,ɪ	u:,ʊ
	e:	o:,ɔ
	ɛ:,ɛ	ə
	æ	ɪ:,ɪ
	/e:/ was probably /i:/ or /ɛ:/ in certain types of pronunciation	
	[a] and [a:] occur as contextual variants of /æ/ and /ɒ:/	
<i>Diphthongs</i>	əi,əu,iu (or ju),eu,ou,ɔi,ui,ei.	
<i>Consonants</i>	p,b,t,d,k,g,ʃ,ʒ	
	m,n,ŋ	
	l,r	
	f,v,θ,ð,s,z,ʒ (later, in medial positions),h	
	j,w (allophone [ɹ] after /h/)	

Text (Macbeth, Act II, Scene 1)

nəu ɔ:r ðə wɪn hæ:f wɪrld
 nɛ:tər si:mz dɛd, ənd wɪkɪd drɛ:mz əbju:z
 ðə kɪrteɪnd sli:p: wɪfkræft sɛlbrɛ:ts
 pɛ:l hɛkəts ɒfəriŋz: ənd wɪðərd mɪrdər,
 əlɑrəmd bæi hɪz sɛntɪnəl, ðə wʊlf,
 hu:z hæulz hɪz wɑf, ðɪrs wɪθ hɪz stɛlθɪ pɛrs,
 wɪθ tɑrkwɪnz rævɪʃŋ strɑɪdz, tu:ərdz hɪz dɪzəɪn
 mu:vz lɛɪk ə gɔ:st. ðəu sju:r⁶ ənd fɛrn-sɛt ɛrθ
 hɛ:r nɒt məi stɛps, hwɪf wɛi ðɛi wɔ:k, fər fɛr:
 ðəi vɛrɪ stɔ:nz prɛ:t əv məi hwɛ:rəbɛut,
 ənd tɛ:k ðə prɛzənt hɔrər frəm ðə tɛɪm,
 hwɪf nəu sju:ts⁶ wɪð ɪt.

6.2.8 The Present English Sound System

<i>Vowels</i>	i:,ɪ	u:,ʊ
	e	
	ɜ:,ə	ɔ:
	æ	
	ʌ	
		ɑ:,ɒ

⁶ Alternatively, [ʃ] or [ʒ] for [sʃ].

<i>Diphthongs</i>	eɪ,əʊ,aɪ,aʊ,ɔɪ,ɪə,ɛə,ʊə
<i>Consonants</i>	p,b,t,d,k,g
	ʃ,ʒ
	m,n,ŋ
	l,r
	f,v,θ,ð,s,z,ʒ,h
	j,w

6.2.9 Modifications in the English Sound System

(1) *Distribution of phonemes.* The similarities of the systems given above may obscure the fact that the same sound, especially as far as the vowels are concerned, may occur in different categories of words according to the period. Thus [u:], now in *food*, occurred in OE in words such as *town*; [i:], now in *team*, occurred in OE in *time*. The following summary shows some of the most striking changes affecting the vowel quality used in particular types of word:

	OE	ME	eModE	PresE
<i>time</i>	i:	i:	əi	aɪ
<i>sweet</i>	e:	e:	i:	i:
<i>clean</i>	æ	ɛ:	e: (or [i:])	i:
<i>stone</i>	ɑ:	ɔ:	o:	əʊ
<i>name</i>	a	a:	ɛ:	eɪ
<i>moon</i>	o:	o:	u:	u:
<i>house</i>	u:	u:	əu	aʊ
<i>love</i>	ʊ	ʊ	ɪ	ʌ

(2) *Vowel changes.* Several trends become apparent from a study of quality changes:

(a) OE long vowels have closed or diphthongized; on the other hand, PresE [əʊ] and [eɪ] show signs of monophthongization.

(b) Certain phonemic qualitative oppositions have coalesced, e.g. OE /e:/ and /æ/; the originally separate diphthongs of *day* and *way*; the diphthong of *know* with the originally pure vowel of *no*; the diphthongs of *day*, *way* with the former pure vowel of *name*; OE /y:,y/ with /i:,ɪ/.

(c) Short vowels, with the notable exceptions of the OE /ɑ,æ/ (and the short diphthong /ɛə/) in open syllables, and ME /ʊ/, have remained relatively stable.

(d) Rounded front vowels have been lost, e.g. OE /y:,y/ and earlier /ø:,ø/.

(e) The loss of post-vocalic [r] in the eighteenth century gave rise to the PresE centring diphthongs /ɪə,ɛə,ʊə/, the pure vowel /ɜ:/ and introduced /ɑ:,ɔ:/ into new categories of words (*cart*, *port*).

(f) Vowels under weak accent increasingly obscured to [ə] or [ɪ], or have been elided.

(g) Changes of quantity have affected certain phonemes in particular contexts or sets of words, e.g. lengthening of OE /ɑ,æ,ɛə/ in open syllables and of ME /ɑ/ + /f,θ,s/; and shortening of ME /o:/ in words like *good*, *book*, *blood*, and of ME /ɛ:/ in such words as *breath*, *death*, *head*.

(3) *Consonant changes.* Changes in the consonantal system are less striking, but the following may be noted:

(a) Double (or long) consonants within words were lost by late ME; certain other consonant clusters ceased to be tolerated, e.g. /hl,hr,hn/ by ME and /kn,gn,wr/ in the eModE period; post-vocalic /r/ was lost in much of the south-east of England in the eighteenth century.

(b) Allophones of certain phonemes have been lost, e.g. the [ɣ] allophone of /g/ in late OE and the [x,ç] allophones of /h/ in eModE.

(c) New phonemes have emerged, e.g. /tʃ dʒ/ in OE, /v,ð,z/ in ME, and /ŋ,ʒ/ in eModE; in addition, /h/ is used initially in words of French origin where, originally, no [h] sound was pronounced (*habit, herb, humble*, etc).

7

Standard and Regional Accents

7.1 *Standards of Pronunciation*

The British are today particularly sensitive to variations in the pronunciation of their language. The 'wrong accent' may still be an impediment to social intercourse or to advancement or entry in certain professions. Such extreme sensitivity is apparently not paralleled in any other country or even in other parts of the English-speaking world. There are those who claim, from an elocution standpoint, that modern speech is becoming increasingly slovenly, full of 'mumbling and mangled vowels and missing consonants'. Alexander Gil and others made the same kind of complaint in the seventeenth century. There is, in fact, no evidence to suggest that the degree of obscuration and elision is markedly greater now than it has been for four centuries. Of more significance—social as well as linguistic—is the attitude which regards a certain set of sound values as more acceptable, even more 'beautiful' than another. Judgements of this kind suggest that there is a standard for comparison; and it is clear that such a standard pronunciation does exist, although it has never been explicitly imposed by any official body. A consideration of the origins and present nature of this unofficial standard goes some way towards explaining the controversies and emotions which it arouses at the present day.

7.2 *The Emergence of a Standard*

It is clear that the controversy does not centre around the written language: the spelling of English was largely fixed in the eighteenth century; the conventions of grammatical forms and constructions as well as of the greater part of our vocabulary have for a long time been accepted and adhered to by the majority of educated English speakers. Indeed, the standardization of the written form of English may be said to have begun in the ninth and tenth centuries. But there has always existed a great diversity in the spoken realizations of our language, in terms of the sounds used in different parts of the country and by different sections of the community. On the one hand, the sounds of the language always being in process of change, there have always been at any one time disparities between the speech sounds of the younger and older generations; the speech of the young is

traditionally characterized by the old as slovenly and debased. On the other hand, especially in those times when communications between regions were poor, it was natural that the speech of all communities should not develop either in the same direction or at the same rate; moreover, different parts of the country might be exposed to different external influences (e.g. foreign invasion) which might influence the phonetic structure of the language in a particular area. English has, therefore, always had its regional pronunciations in the same way that other languages have been pronounced in a variety of ways for basically geographical reasons. Yet, at the same time, especially for the last five centuries, there has existed in this country the notion that one kind of pronunciation of English was socially preferable to others; one regional accent began to acquire social prestige. For reasons of politics, commerce, and the presence of the Court, it was to the pronunciation of the south-east of England and, more particularly, to that of the London region that this prestige was attached. The early phonetician John Hart notes (1569) that it is in the Court and London that 'the flower of the English tongue is used . . . though some would say it were not so, reason would we should grant no less: for that unto these two places, do daily resort from all towns and countries, of the best of all professions, as well of the own landmen, as of aliens and strangers . . .' Puttenham's celebrated advice in the *Arte of English Poesie* (1589) recommends 'the usual speech of the Court, and that of London and the shires lying about London within 60 miles and not much above . . . Northern men, whether they be noblemen or gentlemen, or of their best clerks, [use an English] which is not so courtly or so current as our Southern English is.' Nevertheless, many courtiers continued to use the pronunciation of their own region; we are told, for instance, that Sir Walter Raleigh kept his Devon accent. The speech of the Court, however, phonetically largely that of the London area, increasingly acquired a prestige value and, in time, lost some of the local characteristics of London speech. It may be said to have been finally fixed, as the speech of the ruling class, through the conformist influence of the public schools of the nineteenth century. Moreover, its dissemination as a class pronunciation throughout the country caused it to be recognized as characteristic not so much of a region as of a social stratum. With the spread of education, the situation arose in which an educated man might not belong to the upper classes and might retain his regional characteristics; on the other hand, those eager for social advancement felt obliged to modify their accent in the direction of the social standard. Pronunciation became, therefore, a marker of position in society.

7.3 *The Present-Day Situation: RP*

(1) Some prestige is still attached to this implicitly accepted social standard of pronunciation. Often called RECEIVED PRONUNCIATION (RP), the term suggesting that it is the result of a social judgement rather than of an official decision as to what is 'correct' or 'wrong', it has become more widely known and accepted through the advent of radio and television. The BBC used to recommend this form of pronunciation for its announcers mainly because it was the type which was most widely understood and which excited least prejudice of a regional kind. Indeed, attempts to use announcers who had a mild regional accent used to provoke protests even from the region whose accent was used. Thus, RP often

became identified in the public mind with 'BBC English'. This special position occupied by RP, basically educated southern British English, has led to its being the form of pronunciation most commonly described in books on the phonetics of British English and traditionally taught to foreigners.

(2) Nevertheless, it cannot be said that RP is any longer the exclusive property of a particular social stratum. This change is due partly to the influence of radio and television in constantly bringing the accent to the ears of the whole nation but also, in considerable measure, to the modifications which are taking place in the structure of English society. Just as the sharp divisions between classes have disappeared, so the more marked characteristics of regional speech and, in the London region, the popular forms of pronunciation are tending to be modified in the direction of RP, which is equated with the 'correct' pronunciation of English. This tendency does not mean that regional forms of pronunciation show signs of disappearing; but it has to be recognized that those who wish, for any reason, to modify their speech have models of RP always readily available to their ears while, at the same time, the social inhibitions concerning movement between classes, which were formerly so strongly operative, no longer exert the same pressure.

Moreover, it must be remarked that some members of the present younger generation reject RP because of its association with the 'Establishment' in the same way that they question the validity of other forms of traditional authority. For them, real or assumed regional or popular accent has a greater (and less committed) prestige. It is too early to predict whether such attitudes will have any lasting effect upon the future development of the pronunciation of English. But if this tendency were to become more widespread and permanent, the result could be that, within the next century, RP might be so diluted that it could lose its historic identity, and that a new standard with a wider popular and regional base would emerge. Such a change is made more likely through the recent more permissive attitude of the BBC (and of the commercial television companies) in their choice of announcers, many of whom now have markedly non-RP or non-British accents.

(3) Certain types of regional pronunciation are, indeed, firmly established. Some, especially Scottish English speech, are universally accepted; others, particularly the popular forms of pronunciation used in large towns such as London, Liverpool, or Birmingham, are generally characterized as ugly by those (especially of the older generations) who do not use them. This rejection of certain sounds used in speech is not, of course, a matter of the sounds themselves: thus, [paɪnt] may be acceptable if it means *pint*, but 'ugly' if it means *paint*. It is rather a reflection of the social connotations of speech which, though they have lost some of their force, have by no means disappeared. Indeed, RP itself can be a handicap if used in inappropriate social situations, since it may be taken as a mark of affectation or a desire to emphasize social superiority.¹ It may be said, too, that if improved communications and radio have spread the availability of RP, these same influences have rendered other forms of pronunciation less remote and strange. An American pronunciation of English, for instance, is now completely accepted in Britain; this was not the case at the time when the first sound films were shown in this country, an American pronunciation then being considered strange and even difficult to understand. Speakers of RP are becoming increasingly aware of the fact that their type of

¹ For a summary of experiments on the social evaluation of RP using the matched-guise technique, see Giles *et al.* (1990).

pronunciation is one which is used by only a very small part of the English-speaking world.

(4) Within RP, those habits of pronunciation that are mostly firmly established tend to be regarded as 'correct', whilst innovation tends to be stigmatized. Thus conservative forms tend to be most generally accepted, sometimes even by those who themselves use other pronunciations. Where the accentual patterns or the phonemic structure of words is concerned, this attitude may result in a speaker's use of the conservative variant in a formal situation and the use by the same speaker of a less well-established variant in more casual speech, e.g. the avoidance of /verɪ'fæəbl/ (*verifiable*) and /'dʒuəriŋ/ (*during*) in more formal speech and their replacement with the more conservative /'verɪfəəbl/ and /'dʒuəriŋ/. It may be of interest that the pronunciation /'dʒuəriŋ/ with initial coalescent assimilation was acknowledged by Daniel Jones in the *English Pronouncing Dictionary* in the 1960s and noted as long ago as 1913 by Robert Bridges in his *Tract on English Pronunciation*. Nevertheless, there is still some resistance to accepting such coalescence word-initially in accented syllables.

Where realizational variation (below the level of the phoneme) is affected, most speakers are unaware of their own changing speech patterns. Objections to the use of the glottal stop are often made, its use being popularly associated with Cockney speech, and yet its occurrence as a realization of preconsonantal /t/ is increasingly frequent within the speech of the middle and younger generations of RP speakers (see §9.2.8).

(5) Even within RP there are some areas and many individual words where alternative pronunciations are possible. It is convenient to distinguish three main types of RP: General RP, Refined RP, and Regional RP.² The last two types require some explanation. Refined RP is that type which is commonly considered to be upper-class, and it does indeed seem to be mainly associated in some way with upper-class families and with professions which have traditionally recruited from such families, e.g. officers in the navy and in some regiments. Where formerly it was very common, the number of speakers using Refined RP is increasingly declining. This may be because for many other speakers (both of other types of RP and of regional dialects) a speaker of Refined RP has become a figure of fun, and the type of speech itself is often regarded as affected. (The adjective 'Refined' has been chosen deliberately as having positive overtones for some people and negative overtones for others.) Particular characteristics of Refined RP are the realization of /əʊ/ as [ɛʊ], and a very open word-final /ə/ (and where [ə] forms part of /ɪə, eə, uə/) and /ɪ/. The vowel /ɜ:/ is also pronounced very open, this time in all positions. The vowel /æ/ is often diphthongized as [ɛæ].

While Refined RP reflects a class distinction and describes a type of pronunciation which is relatively homogeneous, Regional RP reflects regional rather than class variation and will vary according to which region is involved in 'regional'. Some phoneticians, on the basis that part of the definition of RP is that it should not tell you where someone comes from, would regard the term 'Regional RP' as a contradiction in terms. Yet it is useful to have such a term as 'Regional RP' to describe the type of speech which is basically RP except for the presence of a few regional characteristics which go unnoticed even by other speakers of RP. For example, vocalization of dark [ɚ] to [ʊ] in words like *held* [heʊd] and *ball* [bɒʊ], a

characteristic of Cockney (and some other regional accents), now passes virtually unnoticed in an otherwise fully RP accent (listen, for example, to umpires at Wimbledon saying *all*.) Or, again, the use of /æ/ instead of /ɑ:/ before voiceless fricatives in words like *after*, *bath*, and *past* (part of the general Northern accent within England) may be likewise acceptable. But some other features of regional accents may be too stigmatized to be acceptable as RP, e.g. realization of /t/ by glottal stop word-medially between vowels, as in *water* (Cockney), the lack of a distinction between /ʌ/ and /ʊ/ (Northern), or the fronting of /u:/ to [y:] (Scottish).

The concept of Regional RP reflects the fact that there is nowadays a far greater tolerance of dialectal variation in all walks of life, although, where RP is the norm, only certain types of regional dilution of RP are acceptable. It remains true, however, that most manuals and dictionaries of the pronunciation of British English, like this book, are based almost entirely on RP.

RP, Refined RP, and Regional RP are not accents with precisely enumerable lists of features but rather represent clusterings of features, such clusterings varying from individual to individual. Thus there are not categorial boundaries between the three types of RP nor between RP and regional pronunciation; a speaker may, for example, generally be an RP speaker but have one noticeable feature of Refined RP.

(6) Finally, it has to be recognized that the role of RP in the English-speaking world has changed very considerably in the last century. Over 300 million people now speak English as a first language, and of this number native RP speakers form only a minute proportion; the majority of English speakers use some form of American pronunciation. However, despite the discrepancy in numbers, RP continues for historical reasons to serve as a model in many parts of the world; and, if a model is used at all, the choice is still effectively between RP and an American pronunciation. When it is a question of teaching English as a second language, there is clearly even greater adherence to one of the two main models. Most teaching textbooks describe either RP or General American, and allegiances to one or the other tend to be traditional or geographical: thus, for instance, European countries continue on the whole to teach RP, whereas some parts of Asia and Latin America follow the American model (see also Chapter 13).

7.4 Comparing Systems of Pronunciation

A comparison of two types of pronunciation will reveal differences of several kinds (as mentioned in §5.3.5):

(a) *Realizational differences*. The system, i.e. the number of distinctive (phonemic) terms operating may be the same, but the PHONETIC REALIZATIONS of the phonemes may be different: e.g. the RP opposition between the vowels of *bet* and *bat* may be maintained, but the realization of both vowels is much more open than in RP (as in Northern English) (see §§8.9.3–4), so that the sound of /æ/ may come near to that of one type of RP /ʌ/ (see §8.9.5); or when, as in Cockney, an allophone [ʔ] represents /t/ between vowels (see §9.2.8); or when the final allophone of /l/ is [ɫ] rather than [ɫ̥] (see §9.7.1).

(b) *Systemic differences* (i.e. differences in PHONEME INVENTORY). The system may

² cf. Wells (1982: 280–3, 297–301).

be different, i.e. the number of oppositions may be smaller or greater, e.g. the RP /æ/-/ɑ:/ opposition may not be present in those Ulster or Scottish forms which do not distinguish *Sam* and *psalm*; or when RP /aɪ/ homophones, as in *side* and *sighed*, are differentiated qualitatively or quantitatively, as in some types of Scottish English; or when the presence of /g/ after [ŋ] in such a word as *sing* deprives [ŋ] of its phonemic status (see §9.6.3).

(c) *Lexical differences* (i.e. differences of LEXICAL INCIDENCE). The system may be the same, but the incidence of phonemes in words is different, e.g. in those Northern forms which have the RP opposition /u:-/ʊ/, but nevertheless use /u:/, in *book*, *took*, etc. (see §8.9.9–10); or when /ɒ/ is used instead of /ʌ/ in *one*, *among*, etc., though the opposition /ɒ/-/ʌ/ exists (see §8.9.5); or when the choice of phoneme is associated with the habits of different generations, e.g. /ɔ:/ for /ɒ/ in *off*, *cloth*, *cross*, etc. (see §8.9.7) or /eɪ/ for /ɪ/ in *Monday*, *holiday*, etc.

(d) *Distributional differences*. The system may be the same, but the phonetic context in which certain phonemes occur may be limited, e.g. in RP /r/ has a LIMITED DISTRIBUTION, being restricted in its occurrence to prevocalic position as in *red* or *horrid*. Accents which display this limited distribution of /r/ are referred to as NON-RHOTIC accents, whilst those in which /r/ has a full distribution (such as most American and Scottish accents) are termed RHOTIC. In the latter accents /r/ occurs pre-consonantly and pre-pausally as well as pre-vocalically; thus *part* and *car* will be pronounced /pɑ:rt/ and /kɑ:r/ whereas in non-rhotic accents the pronunciation will be /pɑ:t/ and /kɑ:/. See §§9.7.2 and 12.4.7.

7.5 Current Changes within RP³

(1) *Realizational changes*. RP /æ/ is frequently heard with a more open quality approaching [a]. This continues a trend in which this RP vowel was typically around [ɛ] early in this century. It appears to conflict with another trend whereby RP /ʌ/ was becoming more fronted and also approaching [a]. There is no evidence suggesting that the two vowels are coalescing; indeed, it seems more likely that /ʌ/ is retreating to its central position.

Other developments among the vowels include /eə/ becoming monophthongal [ɛ:], and /aɪ/ and /aʊ/ having the same centrally open starting-point (as shown in the revision of the first symbol of /aʊ/ from previous editions).

The vowel represented by the spelling <y> at the ends of words like *pity*, *cruelty*, *lengthy* is more and more frequently heard with a closer and more forward pronunciation than the usual realization of /ɪ/, e.g. in *sit*. Indeed the two vowels in such a pronunciation of *city* are far less similar to one another than the two vowels in *meaty*. Thus it seems best to regard the newer, closer, pronunciation as involving unaccented /i:/ (although, of course, theoretically the distinction may be said to be neutralized in this position—see §5.3.4).⁴

As has been mentioned in §7.3 (4), the realization of preconsonantal /t/ as a glottal stop is increasingly common in present-day RP. (See §9.2.8)

(2) *Systemic changes*. The one recent systemic change that is now more or less completed is the loss of /ɔə/ from the phoneme inventory.

(3) *Lexical changes*. There is a strong trend towards selecting /ə/ instead of /ɪ/ in weak syllables, the choice of /ə/ being particularly favoured after /l/ and even more so after /r/, e.g. *angrily* /'æŋgrɪli/ > /'æŋgrəli/. For further detail and examples, see §8.9.2.

Another noticeable trend is the replacement of /ʊə/ by /ɔ:/ in many common words, e.g. *poor* /pɔ:/, *sure* /ʃɔ:/, though /ʊə/ still retains its phonemic status, its contrastive function being illustrated in the speech of most speakers by such sets as *doer*, *dour*, *door* /dʊə, dʊə, dɔ:/.

(4) *Distributional changes*. The most noteworthy trend concerning a regular change in the occurrence of a phoneme is the loss of /j/ after alveolar consonants in such words as *allude* /ə'ljʊ:d/ > /ə'lu:d/, *luminous* /'ljʊ:mɪnəs/ > /'lu:mɪnəs/, *supersede* /sju:pə'si:d/ > /su:pə'si:d/. /j/ is most commonly dropped after /l/ and /s/ (as, indeed, it was long ago after /r/). In sequences of /n/ + /j/, elision of the /j/ is increasingly common in British English. In the case of the alveolar plosives + /j/, coalescence whereby /tj,dj/ > /tʃ,dʒ/, rather than elision, is now increasingly common except initially in an accented syllable, where /t/ + /j/ or /d/ + /j/ tend to be retained. Thus *educate* /'edju:kɪt/ > /'ɛdʒu:kɪt/, *statuesque* /stætju:'esk/ > /stæfju:'esk/.

(5) *Word accent changes*. Certain patterns may be detected, especially in the change affecting adjectives in *-able/-ible* and *-ary/-ory*. In both classes of words, the accent tends now to fall later in the word, thus 'applicable > applicable, 'explicable > explicable, 'justifiable > justifiable, 'fragmentary > fragmentary, 'mandatory > man'datory.

Similarly, the feminine suffix *-ess* increasingly attracts primary accent in words like *coun'tess*, *lio'ness*, *prio'ress*, *stewar'dess*.

Other current changes do not display such regular patterns, and it remains to be seen which of two variant pronunciations at present coexisting will prevail.

7.6 Systems and Standards other than RP

The remainder of this book is a description of English set within the basic framework of RP, with some reference to variation in other dialects in the discussion of each of the RP phonemes. But there are a number of reasons why such particular differences should be drawn together to show the major overall differences between the phonemic system of RP and that of other major dialects of English. In this section we survey briefly differences between RP and five other systems: General American, Scottish English, Northern (England) English, Cockney, and Australian English. We survey an American pronunciation because, as noted in §7.3, this is more frequently the standard model for learners of English as a second language in much of Asia and Latin America. We look at Scottish English because this is the type of pronunciation of English within the British Isles which is most frequently accepted as an alternative standard to RP. We survey Northern (England) English and Cockney because these are the areas (apart from Scottish) whose characteristic pronunciations are heard most widely within Britain and which often underlie regional forms of RP. We look at Australian English because this is typical of an English pronunciation of the southern hemisphere and may increasingly become the standard for a wider area rather than just Australia. Of

³ See further in Ramsaran (1990a).

⁴ For fuller discussion, see Lewis (1990).

course, we could easily have made a case for the inclusion of other systems of pronunciation here (e.g. Caribbean English and Indian English); but since this is not primarily a book about varieties of English, a limit had to be set somewhere. Moreover there are now books which survey dialectal variation in English pronunciation in detail.⁵ Where reference is made in this book to non-standard varieties of English, the type of pronunciation being referred to is the *BASILECTAL* variety of the area concerned, i.e. that used by lower socio-economic classes (and by middle socio-economic classes in informal situations).

7.6.1 General American

The traditional (although not undisputed) division of the United States for pronunciation purposes is into Eastern (including New England and New York City), Southern (stretching from Virginia to Texas and to all points southwards), and General (all the remaining area). General American (GA) can thus be regarded as that form of American which does not have marked regional characteristics (and is in this way comparable to RP). It is the standard model for the pronunciation of English as an L2 in parts of Asia (e.g. the Philippines) and parts of Latin America (e.g. Mexico).

There are two areas of systemic difference between RP and GA. First, GA has no /ɒ/. Most commonly, those vowels which have /ɒ/ in RP are pronounced with /ɑ:/ in GA, e.g. *cod*, *spot*, *pocket*, *bottle*. But a limited subset has /ɔ:/, e.g. *across*, *gone*, *often*, *cough* (as can be seen from the examples, these frequently involve a following voiceless fricative). Secondly, GA lacks the RP diphthongs /ɪə, eə, uə/ which correspond in GA to sequences of vowel plus /r/, e.g. *beard*, *fare*, *dour*, /bɪrd/, /fɛr/, /dʊr/. This reflects the allied distributional difference between RP and GA, namely that, unlike RP, where /r/ occurs only before vowels, GA /r/ can occur before consonants and before pause (GA is called a rhotic dialect and RP a non-rhotic dialect).

The main difference of lexical incidence concerns words which in RP have /ɑ:/ while in GA they have /æ/. Like the change from /ɒ/ to /ɔ:/, this change commonly involves the context before a voiceless fricative, or alternatively before a nasal followed by another consonant; thus RP /pɑ:st/ - GA /pæst/, RP /ɑ:ftə/ - GA /æftə/, RP /pɑ:θ/ - GA /pæθ/.

Differences of realization are always numerous between any two systems of English pronunciation, and only the most salient will be mentioned. Among the vowels this includes the realization of the diphthongs /eɪ/ and /əʊ/ as monophthongs [e:] and [o:], hence *late* [le:t] and *load* [lo:d]. Among the consonants, /r/ is either phonetically [ɹ], i.e. the tip of the tongue is curled further backwards than in RP, or else a similar auditory effect is achieved by bunching the body of the tongue upwards and backwards; /t/ intervocalically is usually a voiced tap in GA, e.g. *better* ['berə]; and /l/ is generally a dark [ɫ] in all positions in GA, unlike RP, where it is a clear [l] before vowels and a dark [ɫ] in other positions (see §9.7.1).

7.6.2 Scottish English

The typical vowel system of Scottish English (SE) involves the loss of the RP distinctions between /ɑ:/ and /æ/, between /u:/ and /ʊ/, and between /ɔ:/ and

/ɒ/. Thus SE pronounces the pairs *ant* and *aunt*, *soot* and *suit*, *caught* and *cot* similarly. SE also has no /ɪə, eə, uə/ because, like American English, it is rhotic, and *beard*, *fare*, and *dour* are pronounced as /bɪrd/, /fɛr/, and /dʊr/. However, the vowel in /fɛr/, which we have transcribed with the RP diphthong /eɪ/, is typically monophthongal [e:] (and of course would be transcribed as such if we were devising a phonemic transcription independently for SE). The vowel /əʊ/ is also monophthongal [o:] as in *coat* [ko:t]; so the vowels in *fare* and *coat* are similar to those in American English. Moreover, the vowel in *soot* and *suit* is not like either of the RP vowels in these words, but is considerably fronted to something like [y], hence [syt].

The chief differences from RP in the realization of the consonants lies in the use of a tap [ɾ], e.g. *red* [red] and *trip* [trɪp], though there is variation between this and [ɹ] (the usual type in RP), the use of [ɹ] being generally more prestigious. The phoneme /l/ is most commonly a dark [ɫ] in all positions, *little* [lɪtɫ], and *plough* [pɫəʊ]. Finally, intervocalic /t/ is often realized as a glottal stop, e.g. *butter* [ˈbʌʔə].

7.6.3 Cockney

We use the term Cockney, rather than London English, because, unlike General American and Scottish English, Cockney is as much a class dialect as a regional one. In its broadest form the dialect of Cockney includes a considerable vocabulary of its own, including rhyming slang. The characteristics of Cockney pronunciation are spread more widely through the working class of London than is its vocabulary. Moreover, some traces of Cockney pronunciation are often present in most middle-class speech of the area.

Unlike the previous two types of pronunciation, there are no differences in the inventory of vowel phonemes between RP and Cockney, and there are relatively few (compared with GA and SE) differences of lexical incidence. There are, however, a large number of differences of realization. The short front vowels tend to be uniformly closer than in RP, e.g. in *sat*, *set*, and *sit*, so much so that *sat* may sound like *set* and *set* itself like *sit* to speakers from other regions. Additionally the short vowel /ʌ/ moves forward to almost C[a]. Among the long vowels, most noticeable is the diphthongization of /i:/ (= [iɪ]), /u:/ (= [uɪ]), and /ɔ:/, which varies between [ɔɪ] morpheme-medially and [ɔwə] morpheme-finally, thus *bead* [bɪd], *boot* [buɪt], *sword* [sɔɪd], *saw* [sɔwə]. Cockney also uses distinctive pronunciations of a number of diphthongs: /eɪ/ = [aɪ], /aɪ/ = [aɪ], /əʊ/ = [æʊ], and /aʊ/ = [a:], e.g. *late* [laɪt], *light* [laɪt], *load* [læʊd], *loud* [la:d]. The last two vowels are close enough to cause considerable confusion among non-Cockney listeners, although the distinction is never actually neutralized.

Among the consonants, most notable are the omission of /h/ and the replacement of /θ, ð/ by /f, v/, e.g. *think* /fɪŋk/, *father* /ˈfʌvə/, *hammer* /ˈæmɜv/. Dark [ɫ], i.e. /l/ in positions not immediately before vowels, becomes vocalic [ʊ], e.g. *milk* [mɪʊk]; /t/ is realized as a glottal stop between vowels, e.g. [bʌʔə] and there is glottal replacement of [p, t, k] before a following consonant, e.g. *soapbox* [ˈsəʊʔbɒks], *statement* [ˈstɑɪʔmənt] *technical* [ˈteʔnɪkəl], as in some types of Scottish English; and /j/ is elided after alveolar plosives, e.g. *student*, *during*.

Cockney has consistently had a major influence on the development of RP, and

⁵ In particular Wells (1982).

nowadays that type of Regional RP which is heavily influenced by Cockney is often referred to as Estuary English (i.e. a middle-class pronunciation typical of the Thames estuary). Particularly characteristic of this type of Regional RP are the replacement of [p,t,k] by [ʔ] before a consonant (see §9.2.8 (b) (ii) below) and the use of [ʊ] in place of [ɪ].

7.6.4 Northern English

While there is relative homogeneity in a broad Cockney accent but much less so in General American and Scottish English, the label 'Northern English' is even less homogeneous. We use it here simply to identify those things which the disparate pronunciations systems in the north of England have in common (and we will also mention a few characteristics which are typical only of certain areas). The area we are talking about covers that area north of a line from the River Severn to the Wash, and includes Birmingham.

The major identifying feature of this area is the loss of the distinction between RP /ʊ/ and /ʌ/, the single phoneme doing duty here varying in quality from [ʊ] to [ʌ]. So Northern English has no distinction between *put* and *putt*, *could* and *cuɔd*, and, for many speakers, between *buck* and *book* (although others may use /u:/ in the latter word). Hypercorrections are often made by those attempting Regional RP producing, for example, *sugar* [ʃʌgə], *pussy* [pʌsɪ], *put* [pʌt]. Almost as identifying a characteristic is the changeover in lexical incidence from /ɑ:/ to /æ/ in words with a following voiceless fricative (or a nasal followed by a further consonant, as in General American), e.g. *past* /pæst/, *laugh* /læf/, *aunt* /ænt/. Another type of lexical incidence concerns the occurrence of a full vowel in prefixes where RP has /ə/, e.g. *advance* /æd'væns/, *consume* /kən'sju:m/, *observe* /ɒb'zɜ:v/. The short vowels are generally realized with more open qualities than RP, e.g. *mad* [mad], and the diphthongs /eɪ/ and /əʊ/ are commonly monophthongal [e:] and [o:] as in GA and SE (indeed sometimes, as in Newcastle, the direction of the diphthong is reversed to [eə] and [oə]). Other vowel changes (compared with RP) characteristic of particular areas include the loss of the /eə/-/ɜ:/ distinction in Liverpool (the local accent is called Scouse and its common realization as [œ:], e.g. both *fare* and *fur* are pronounced [fœ:]; the realization of /aʊ/ as [u:] in Newcastle (where the broad local accent is called Geordie) while /u:/ itself becomes [ɪə], e.g. *about* [əbu:t], *boot* [bɪət]; and the use of a particularly close /ɪ/ in Birmingham, e.g. *pit* is almost [pɪt], where the distinction between *pit* and *peat* will depend on length alone.

Most notable among the consonants of Northern English is the realization of /r/ as [ɹ] in a number of conurbations including Leeds, Liverpool, and Newcastle, and the lack of the RP allophonic difference between clear [l] and dark [ɫ], clear [l] being used in all positions in many areas, e.g. Newcastle, and dark [ɫ] in others, e.g. Manchester. In a quite extensive area, from Birmingham to Manchester and Liverpool, the RP single consonant /ŋ/ becomes /ŋg/, e.g. *singing* [sɪŋgɪŋg]. Also in a number of urban areas, notably south-east Lancashire, /p,t,k/ in final position (i.e. before a pause) are realized as ejectives.

7.6.5 Australian English

There is little regional variation in Australian English (ANE), the variation which does occur being largely correlated with social class and ranging from a broad accent all the way up to regional RP. The broad accent described here shares many features with Cockney, but has of course a particular combination of these and other features which identify it.

As in Cockney, there are no differences of phonemic inventory from RP and no extensive classes of word involved in differences of incidence. It is the realization of long /ɑ:/ as [a:] which more than any other identifies ANE, e.g. *father* [fa:ðə], *part* [pa:t]. As in Cockney, /i:/ and /u:/ are realized as [iɪ] and [uɪ] and the short front vowels are all closer than RP, although [ɪ] does not occur in unaccented positions, being replaced by /ɪr/ word-finally and by /ə/ in other positions, e.g. *city* /səti:/. In its diphthongs ANE is again like Cockney in having /eɪ/ = [aɪ] and /aɪ/ = [aɪ], and in having a convergence of quality of /əʊ/ and /aʊ/; however, diphthongs in /ə/ are monophthongized, so /ɪə/ = [ɪ:], *clear* [klɪ:] (leading to an accumulation of three vowels, /i:/, /ɪ:/, and [ɪ:] in the close front area), /eə/ = [e:], *fare* [fɛ:], while /ʊə/ is either replaced by /ɜ:/ as in *sure* or becomes disyllabic as in *sewer* /su:ə/.

Although ANE does drop /h/, it does not use glottal stop, nor does it vocalize /l/, having dark [ɫ] in all positions.

A particular development in Australian English (and in New Zealand) which has been the subject of much discussion recently, both in newspapers and in academic journals,⁶ is the increasing use of a high rising tone on declarative clauses (where a fall would normally have been expected). The meaning of this tone and the reasons behind its increased use have also been much discussed (see further under §11.6.3).