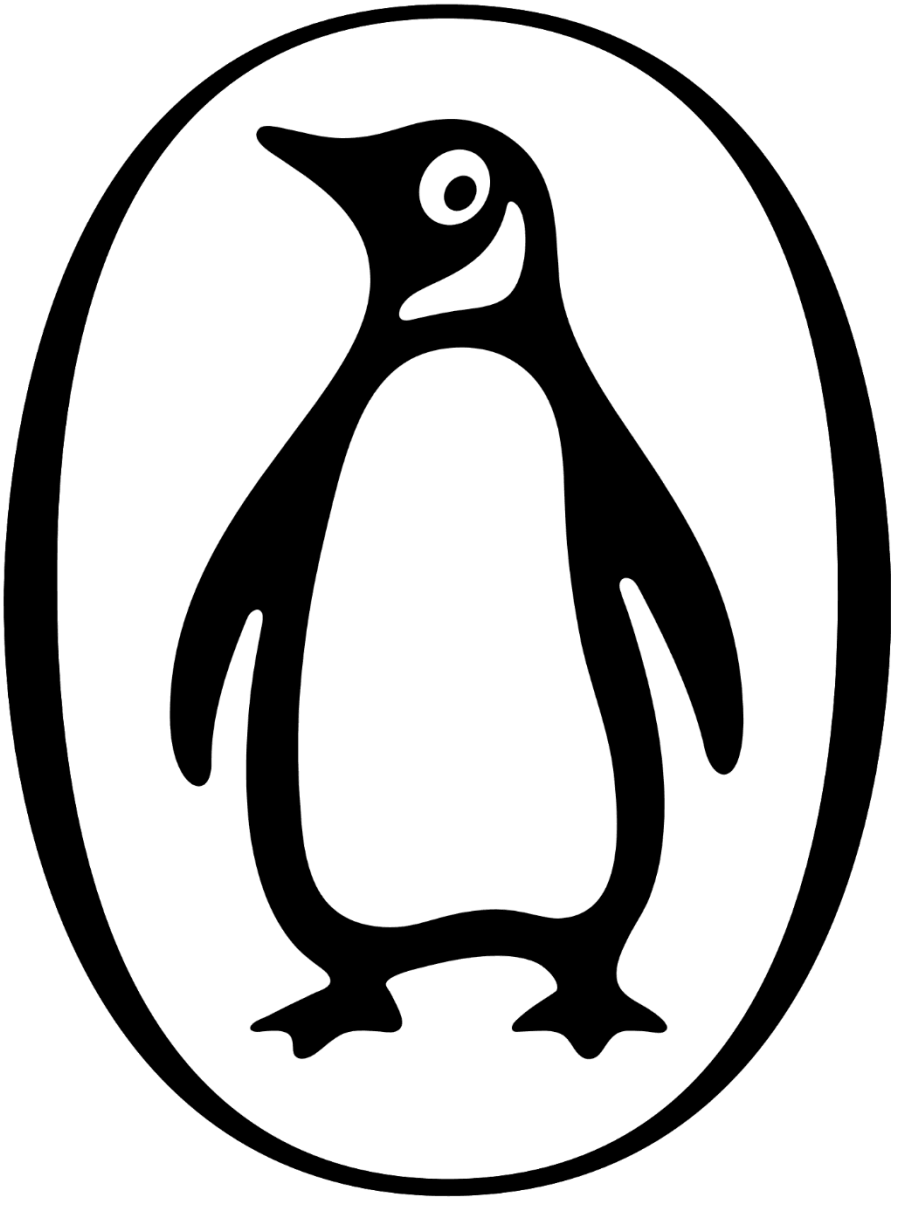

PENGUIN BOOKS

Published by the Penguin Group
Penguin Books Ltd, 80 Strand, London WC2R 0RL, England
Penguin Group(USA) Inc., 375 Hudson Street, New York, New York 10014, USA
Penguin Group(Canada), 90 Eglinton Avenue East, Suite 700, Toronto, Ontario, Canada M4P 2Y3
(a division of Pearson Penguin Canada Inc.)
Penguin Ireland, 25 St Stephen's Green, Dublin 2, Ireland
(a division of Penguin Books Ltd)
Penguin Group(Australia), 250 Camberwell Road, Camberwell, Victoria 3124, Australia
(a division of Pearson Australia GroupPty Ltd)
Penguin Books India Pvt Ltd, 11 Community Centre, Panchsheel Park, New Delhi – 110 017, India
Penguin Group(NZ), cnr Airborne and Rosedale Roads, Albany, Auckland 1310, New Zealand
(a division of Pearson New Zealand Ltd)
Penguin Books (South Africa) (Pty) Ltd, 24 Sturdee Avenue, Rosebank, Johannesburg 2196, South Africa
Penguin Books Ltd, Registered Offices: 80 Strand, London WC2R 0RL, England

www.penguin.com

First published 2005
1

Contents

Discourse

Dialects

Languages

List of Figures

19 The Indo-European family of languages

very divergent – about the way language can or should be studied. This is above all a personal account. Nor is there much on the methodology of linguistic enquiry: while I talk a lot about child language, for example, I do not say anything about the range of methods that linguists use when they are investigating child language. I have therefore included a short bibliography of further reading for those who reach §73 and wish to take such things further.

David Crystal

Holyhead, April 2005

1
How what works?

'Language', the title of this book says. But what is meant by 'language'? Consider the following expressions:

body language

spoken language

written language

sign language

computer language

the French language

bad language

animal language

the language of birds

the language of cinema

the language of music

the language of love

Plainly the word is being used in many ways – some technical, some figurative – and the senses go in various directions. If a reviewer were to remark, after an impressive orchestral concert, 'The conductor and the musicians were all speaking the same language', we would interpret this to be a comment about their playing, not their chatting. And the same point applies to other linguistic terms, when used in special settings. I have seen books called *The Grammar of Cooking* and *The Syntax of Sex*. The first was a collection of recipes – as was the second.

How Language Works is not about music, or cookery, or sex. But it is about how we *talk* about music, cookery, and sex – or, indeed, about anything at all. And it is also about how we write about these things, and send electronic messages about them, and on occasion use manual signs to communicate them. The operative word is 'how'. It is commonplace to see a remarkable special effect on a television screen and react by exclaiming 'How did they do that?' It is not quite so usual to exclaim when we observe someone speaking, listening, reading, writing, or signing. And yet if anything is worthy of exclamation, it is the human ability to speak, listen, read, write, and sign.

An alien visitor to Earth might well wonder what was going on. It would see humans approach each other, use their mouths to exchange a series of noises, and – apparently as a result of making these noises – cooperate in some activity. It would see human eyes look at a set of marks inscribed on a surface, and the eye-owners then behaving in the same way – going out of one door rather than another in a theatre, for instance. Rather less often, it would see some humans using their hands and faces to achieve the same results that others obtain through the use of their mouths. In each case it might think: 'How did they do that?' And in each case the answer would be the same: 'through the use of language'.

But our alien would also observe other kinds of behaviour. It would see humans smiling and frowning at each other, or waving and gesturing, or stroking and kissing. It would notice that the effect of carrying out these actions was similar in some respects to that produced by the use of spoken noises, written marks, and manual signs. It might well reflect: can these actions therefore be called 'language' too?

Our alien would also see apparently similar behaviour among other species. It would see a bee find a source of nectar, return to a hive, and perform a series of dance-like body movements. Other bees would then move off in the direction of the nectar. Animals of all kinds would seem to be sending information to each other in analogous ways. Is this the same sort of behaviour as the humans are displaying, our alien observer might think? Do animals also have language?

These questions involve more than hypothetical extraterrestrials. Terrestrial observers also need to be able to answer them, as a preliminary stage in the study of language. If we pick up a manual called *How Cars Work*, we do not expect to find in it chapters on bicycles and lawn-mowers. Nor, in *How Language Works*, will there be much space devoted to the use of facial expressions and body movements or to the way animals communicate. Why not?

Modes of communication

Because not all of these forms of communication are *language*, in the sense of this book. *Communication* is a much broader concept, involving the transmission and reception of any kind of information between any kind of life. It is a huge domain of enquiry, dealing with patterned human and animal communication in all its modes. Those who study behaviour usually call this domain *semiotics*. *Linguistics*, the science of language, is just one branch of semiotics.

There are five modes of human communication, because there are only five human senses which can act as channels of information: sound, sight, touch, smell, and taste. Of course, if you believe in telepathy, you would need to recognize a 'sixth sense' available for communication; and perhaps there are life

2
How to treat body language

When people talk about body language (§1), they are referring to those features of bodily behaviour which are under some degree of conscious control, and which they can therefore use to express different sorts of meaning. The meanings involved are all fairly 'primitive' expressions of attitude or social relationship, such as affection, aggression, sexual attraction, greeting, congratulation, gratitude, surprise, and the signalling of attention. Both tactile and visual modes of communication are employed.

The tactile mode

The tactile mode of nonverbal communication operates when parts of the body make planned physical contact with other people. A very wide range of meaningful activities is expressed by such contact, as this small selection of terms suggests:

> dig, embrace, hold, jog, kick, kiss, nudge, nuzzle,
> pat, pinch, punch, shake, slap, spank, tickle

They operate within a complex system of social constraints. Some of the acts tend to be found only in private – notably, sexual touching. Some are specialized in function: examples include the tactile activities which we permit from doctors, dentists, hairdressers, or tailors. And some are restricted to certain ceremonies or occasions – a handshake which signals a formal agreement, for example, or a laying on of hands in the context of religion or healing.

The communicative value of tactile activities is usually fairly clear within a culture, but there are many differences

across cultures. Some societies are much more tolerant of touching than others, so much so that a distinction has been proposed between *contact* and *non-contact* societies – those that favour touching (such as Arabs and Latin Americans), and those that avoid it (such as North Europeans and Indians). In some cultures, conversationalists touch each other two or three times a minute; in others, no touching takes place at all.

Related to our use of body contact is the way we use body distance and orientation to communicate meaning. There are norms of proximity within a culture (*distance zones)* which can inform us about the social relationship between the participants. Latin Americans, for example, prefer to stand much closer to each other during a conversation than do North Europeans. In a traditional caste system, such as in India, the acceptable distance zones between the members of different castes can vary greatly – from less than 2 metres to over 20.

The visual mode

We use the visual mode to communicate nonverbally in several ways. We can gesture, vary our facial expressions, make eye contact, and alter our body posture. Each of these behaviours performs a variety of functions. Movements of the face and body give clues to our personality and emotional state. The face, in particular, signals a wide range of emotions, such as fear, happiness, sadness, anger, surprise, interest, and disgust. Many of the expressions vary in meaning across cultures, and we have to learn how to interpret the sometimes very subtle movements in the faces of people whose racial characteristics differ from our own.

In addition, the face and body send signals about the way a social interaction is proceeding. We use *eye contact* to show

who is the focus of our communication, in a group, or to prompt a person to speak next. We use *facial expressions* to give feedback to others about how we are receiving their message, expressing such meanings as puzzlement or disbelief. We use our *body posture* to convey our attitude towards an interaction – for instance, whether we are interested or bored. Several kinds of social context are associated with specific facial or body behaviours, such as waving upon meeting or taking leave. Ritual or official occasions are often associated with gesture and posture – as with kneeling, standing, bowing, and blessing.

Some visual effects are widely used in the cultures of the world. An example is the *eyebrow flash*, used unconsciously when people approach each other and wish to show that they are ready to make social contact. Each person performs a single upward movement of the eyebrows, keeping them raised for about a sixth of a second. The effect is so automatic that we are hardly ever conscious of it. But we become uneasy if we do not receive an eyebrow flash when we expect one (from someone we know); and to receive an eyebrow flash from someone we do not know can be uncomfortable, embarrassing, or even threatening.

Most gestures and facial expressions, however, differ across cultures. Sometimes the differences are very noticeable, especially when we visit a society which uses far more gestures and facial expressions than we are used to (e.g. Italian, compared with British) or far fewer (e.g. Japanese, compared with British). We even coin phrases to express our sense of these differences, as when an English person describes Italians as 'talking with their arms' or Westerners refer to people from oriental countries as 'inscrutable'.

Even when a visual effect seems to be shared between

societies, we have to be careful, for it can convey very different meanings. A thumbs-up sign has a positive 'all is well' or 'I am winning' meaning in Western Europe, the USA, and other cultures influenced by its use as a symbol of combat survival in Roman times. But in the Arab world, as well as in parts of West Africa and Asia, it is a symbol of insult, equivalent to giving someone 'the finger' ('up yours!') in the West. As a consequence, it was never entirely clear, during the aftermath of the Iraq War of 2003, when Iraqis were seen on television giving a thumbs-up to American troops, whether this was the traditional gesture being used as an insult or whether it was the Western version being adopted as a sign of cooperation and a symbol of freedom.

Conversely, a particular meaning can be conveyed by a variety of different visual signals. To express humility or deference, for example, Europeans tend to extend or lower their arms, and they sometimes bow their heads. But in other cultures we find more profound bowing, using the whole of the upper half of the body, as well as crouching, crawling, and prostration. We also see other kinds of hand or arm movement, such as the placing of the palms together in an upward orientation (as in the Indian subcontinent).

Properties of language

Body language is evidently an important means of human communication, and when it comes to basic emotions and social relationships, it is a familiar experience that a gesture, facial expression, or piece of bodily contact can 'speak louder' than words. However, the potential of body language to express meaning is very limited, compared to that which is made available through speaking, writing, or signing.

such as bee-dancing or birdsong, are highly limited in what they can do, compared with language. Bees can 'talk' about nectar but not about much else.

There are several other important differences between animal communication and language. In particular, language enables us to talk about events remote in space or time from the situation of the speaker: I can talk about what happened in the near or remote past and speculate about the near or remote future. This property of language – often called *displacement* – is something which goes well beyond the capabilities of animal signals, which reflect stimuli (such as the presence of danger or the direction of a food source) encountered in the animal's immediate environment.

Despite some superficial similarities, so-called 'body language' and 'animal language' are very different from what happens in language, in the sense of this book. I find it clearer to avoid the use of the term *language* altogether, in fact, and to describe these phenomena in more general terms – as *body communication* and *animal communication*. There is nothing wrong with the 'language' metaphor, of course, as long as we realize that that is what it is – no more than a vague approximation to the structurally complex and multifunctional behaviour we find whenever we speak, write, or sign.

3
How we use the ‘edges’ of language

The contrast between what counts as language and what does not is usually clear enough, once we look for evidence of productivity and duality of structure in communicative behaviour (§2). But the boundary is fuzzy at times. In particular, some non-linguistic forms of behaviour, both vocal and visual, can be adapted so that they take on some of the functions of language. There are also some features of language which are decidedly less complex than others, and where it is unclear whether they should count as part of language or not.

Making vocal noises

The vocal organs (§4) can be used to make a wide range of noises that are definitely not linguistic. They express only a biological state, and communicate no cultural meaning. Examples are coughing, sneezing, and snoring, as well as the various voice qualities which signal a physical condition, such as hoarseness. It makes no sense to talk about ‘snoring in English’, nor do we expect a foreign language course to teach us how to cough. But a phenomenon such as whistling does something more than just express a basic biological or psychological state.

When we blow air through tensed and rounded lips, we form a primitive musical instrument, and a note is the result. We can alter the pitch level by moving the tongue and cheeks to change the shape of the inside of the mouth. We can alter its loudness by blowing harder. And we can alter its quality (making it soft or shrill) by altering the tension of the mouth

muscles or putting our fingers against the lips to make the sound sharper. It does not come naturally. Children have to learn to whistle. And the behaviour is subject to social factors: usually it is boys and men who whistle.

People most often use whistling to carry musical melodies, and some professional whistlers have developed this musical skill into an art form. Some people are able to mimic birdsong or other animal noises. But these imitative abilities have no more linguistic significance than the phenomena they imitate. We move in the direction of language only when individual whistles are used conventionally in a culture to express a specific and shared meaning. Examples include the 'wolf-whistle', the whistle which calls sharply for attention, the whistle of amazement, and the whistle of empathy ('gosh!'), which is often more a breathy exhalation than an actual whistle.

We see the communicative potential of whistling at its most developed in the case of the so-called *whistle languages*, found in some Central and South American tribes, as well as in the occasional European community, such as in the Pyrenees, Turkey, and the Canary Islands. Conversations have been observed between people standing at a considerable distance from each other, especially in mountainous areas, carried on entirely in whistles. The whistled conversations deal with quite sophisticated and precise matters, such as arranging a meeting or selling some goods.

Whistled speech closely corresponds to the tonal and rhythmical patterns of spoken language, and is especially complex when the whistlers speak a language in which pitch levels (tones) are important, such as some languages of Central America. With very few exceptions, each 'syllable' of whistle

corresponds to a syllable of speech. Ambiguity is uncommon, because the topic of the conversation is usually something evident in the situation of the speakers. However, it is important for both speakers to use the same musical key, otherwise confusion may arise.

Whistled dialogues tend to contain a small number of exchanges, and the utterances are short. They are most commonly heard when people are at a distance from each other (e.g. when working the land), but they can also be found in a variety of informal settings. Although women are able to understand whistled speech, it is normally used only by and between men.

The whistling is a substitute for speech. The whistlers stop whistling when they are within normal speaking range of each other, and talk in the normal way. For this reason, whistled speech has been called a speech replacement, or *surrogate*. Because it is used only in certain circumstances, and can convey only a limited range of meanings, and is not used equally by all members of the community, it does not really correspond to the complexity and functional breadth of a spoken language. But it is certainly a step in that direction.

Being paralinguistic

There are a number of vocal noises which can be 'superimposed' on the stream of speech. It is possible, for example, to speak while sobbing, crying, laughing, or giggling, or we can introduce a tremulous 'catch' in our voice. The auditory effects are usually immediate and dramatic – though the meaning conveyed is sometimes unclear, and always subject to cultural variation. A giggle can convey humour, innuendo, sexual interest, and several other nuances. In Britain

communication in particular situations. They are often referred to as *sign languages*, but few have developed any degree of structural complexity or communicative range, and it is therefore important to distinguish them from sign language proper – the natural signing behaviour which has evolved for use among deaf people (§25). Nonetheless, they are a definite advance on the 'basic' kinds of body language that are seen in everyday interaction (§2).

Several professions use sets of conventional signals. Sports players and officials can use hand and arm gestures to show the state of play, or an intention to act in a certain way. Groups of performers (such as acrobats or musicians) use them to coordinate their activities. In casinos, officials use them to report on the way a game is going, or to indicate problems that might affect the participants. In theatres and cinemas, ushers use them to show the number and location of seats. In sales rooms, auctioneers use them to convey the type and amount of selling and buying.

People controlling cranes, hoists, and other equipment can signal the direction and extent of movement. In aviation marshalling, ground staff can send visual information about the position of an aircraft, the state of its engines, and its desired position. Firemen can send directions about the supply of water, water pressures, and the use of equipment. Divers can communicate depth, direction, and time, and the nature of any difficulties they have encountered. Truck drivers can exchange courtesy signals, give information about the state of the road, or show they are in trouble. Environmental noise may make verbal communication impossible (e.g. in cotton mills), so that workers start signing to each other.

Race-course bookies send hand and arm signals about the

number of a race or horse, and its price. In radio and television production, producers and directors can signal to performers the amount of time available, instructions about level of loudness or speed of speaking, and information about faults and corrections. Religious or quasi-religious groups and secret societies often develop ritual signing systems so that members can recognize and communicate with each other. Some monastic orders have developed signing systems of considerable sophistication, especially if their members are vowed to silence, as in the case of the Trappist monks.

But in none of these cases are we dealing with systems containing thousands of expressive possibilities, as we are with 'language proper', or with signed sentences of any complexity. They all lack productivity and duality of structure (§2), and they are meaningless outside of the situations for which they were devised. A Trappist monk would make little headway signing at a football referee, and vice versa. These signalling systems are highly restricted methods of communication, invented to solve a particular problem. They are a step or so above basic body gestures, but not much more than that.

There are only a few cases where the visual and tactile modes have been adapted to perform a truly linguistic function, providing alternative modes of communication to that which we encounter most commonly in the auditory-vocal behaviour we call 'speech'. Most obviously, the visual mode is used in writing, and there have been several writing-based visual codes, such as semaphore and Morse. Writing can involve a tactile dimension, too, as when visually impaired people receive written information through their finger-tip contact with the sets of raised dots known as braille. Tactile codes also exist, in which sounds or (more usually) letters are

communicated through touch; and touch is critical when deaf-blind people use their hands to sense the movements of another person's vocal organs while speaking. Finally, facial expression and gesture is crucially involved in deaf signing – which, as we shall see (§25), is very different from the signing systems described above or the everyday gestural behaviour used by hearing people.

Speech, writing, sign: these are the three mediums which define the conceptual domain of any book on language. And the natural place to begin is with speech. Of all the modes of communication (§1), the auditory-vocal medium is the one which has been most widely adapted for purposes of human communication. All children with their sensory and mental faculties intact learn to speak before they learn to write. Only in the case of a child born deaf is there a natural opportunity to learn a non-auditory system. Moreover, all languages exist in spoken form before they are written down. Indeed, some 40% of human languages (over 2,000 in all) have never been written down. For their speakers, the topic 'how language works' could mean only one thing: 'how speech works'.

4
How we make speech sounds: phase 1

The parts of the body used in the production of speech are called the *vocal organs*, and there are more of them than we might expect. We have to take into account the *lungs*, the *throat*, the *mouth*, and the *nose*. Inside the mouth, we must distinguish the *lips*, the *tongue*, the *teeth*, the roof of the mouth (or *palate*), and the small fleshy appendage hanging down at the very back of the palate (the *uvula*). Inside the throat, we find the upper part, or *pharynx*, operating in a different way from the lower part, or *larynx*. And within the larynx (§5) we need to recognize the important role of the *vocal folds*, located behind the Adam's apple. The space between the vocal folds is known as the *glottis*. There is a lot going on at the same time, when we speak.

The pharynx, mouth, and nose form a system of hollow areas, or cavities, known as the *vocal tract*. (Some speech scientists include the larynx and lungs under this heading as well.) When we move the organs in the vocal tract, we alter its shape, and it is this which enables the many different sounds of spoken language to be produced. In fact, it is the remarkable versatility of the human vocal tract which is so noticeable when we compare humans with their nearest animal cousins, the primates.

The primate vocal tract is very different from that found in humans. Primates have long, flat, thin tongues, which have less room to move. Their larynx is higher (§5), and there is little sign of a pharynx. They are

we are engaged in normal conversation – though the amount increases to some extent if our speech becomes loud or effortful, as in shouting, acting, singing, public speaking, or producing a 'stage whisper'.

Lung air is technically called *pulmonic* air (from the Latin word for 'lung', *pulmo*), and when pulmonic air flows outwards, it is said to be *egressive*. The vast majority of speech sounds are made using pulmonic egressive air. It is also possible – though not usual – to speak while the air-stream is flowing inwards to the lungs, as we inhale. This would be called pulmonic *ingressive* air. We do occasionally hear this air-stream used when someone is trying to talk while laughing or crying, or when out of breath. Words such as *yes* and *no* are sometimes said with an ingressive air-stream, when we use a 'routine' tone of voice to acknowledge what someone is saying. An alternate use of egressive and ingressive airstreams is sometimes heard when people are counting rapidly, 'under their breath'. But ingressive speech is of poor quality, muffled, and croaky, and many people find it unpleasant to listen to. It is never put to routine use in everyday speech.

The sequence of events involved when we breathe in and out is known as the *respiratory cycle*. Normally, the two halves of this cycle are nearly equal in duration; but when we speak, the pattern changes to one of very rapid inhalation and very slow exhalation. The rate at which we breathe also changes. When we are silent and at rest, our average rate is 12 breaths a minute, so the time we take to inhale and exhale is about 2.5 seconds each. During speech, we cut down the time for inhaling to as little as a quarter of a second, and we regularly extend the time for exhaling to 5 or 10 seconds or more, depending on our voice control, emotional state, and other

such factors. This altered pattern of breathing enables our exhalations to 'carry' much larger amounts of speech than would otherwise be the case. In everyday conversation, it is perfectly normal to produce 250 to 300 syllables in a minute.

Not using the lungs

The vowels and consonants of English, as of most languages, are all made using pulmonic egressive air. But there are a few types of speech sound which do not use an air-stream from the lungs, and these are encountered in many languages of the world.

Probably the most distinctive type of non-pulmonic sound is the *click*. Click sounds are sharp suction noises made by the tongue or lips. For example, the noise we write as *tut tut* (or *tsk tsk*) is a pair of click sounds, made by the tongue against the top teeth (*dental* clicks). In English they are used to express disapproval. Throughout the Near East a single such click is widely used to express negation. Another type of click uses the side of the tongue (a *lateral* click), made as a noise of encouragement – usually just to horses or other animals, but occasionally to other human beings. A click sound made with the lips puckered is known as a *bilabial* click – it is often used as a 'kiss at a distance'. In each of these cases, we can breathe in and out, quite independently, while making the click sounds. This shows that the lungs are not involved in their production.

In European languages, isolated click sounds are often heard as meaningful noises, but they are not part of their system of vowels and consonants. However, in many other languages, clicks *are* used as consonants. Best known are some of the languages of southern Africa, often referred to as *click*

languages. The Khoisan languages, which include the languages of the Khoikhoin and San tribes, have the most complex click systems, using many different places of articulation in the mouth, and involving the simultaneous use of other sounds made in the throat or nose. There are no less than forty-eight distinct clicks in one such language, !Xũ (the '!' in the name represents one such click).

Although we think of clicks as relatively 'quiet' sounds, they can be pronounced with considerable force. Miriam Makeba's 'click songs' were very popular in the 1960s. A native speaker of Xhosa, she used several words containing click consonants in her singing, achieving notable effects by articulating them with great resonance.

Some languages use other types of non-pulmonic sound. We can use the space between the vocal folds, the *glottis*, to start an air-stream moving. A number of languages make use of sounds based on this principle, referred to as the *glottalic* air-stream mechanism. When the glottis makes the air move inwards, the sounds are called *implosives*. Implosive consonants occur in many languages, but are particularly common in Native American and African languages. To European ears the effect is of the sounds being 'swallowed'.

When the glottis makes the air move outwards, the sounds are called *ejectives*. Ejective consonants are widely used in the languages of the Caucasian family, and also in many Native American and African languages. They may even be heard in certain accents and styles of English. Speakers from the north of England quite often use them at the ends of words, in place of the usual [p], [t], or [k].* And regardless of the accent we use, if we speak in a tense, clipped manner, the effect is one of the sounds being 'popped out' at the end of a word.

Making unusual sounds

In special circumstances, people can speak using an abnormal air-stream mechanism. It is possible to compress air within the cheek-space and use that to carry speech – the so-called *buccal* voice. This is best known through the voice of Walt Disney's Donald Duck. It is also possible to make sounds using air rising from the stomach or oesophagus (the pipe leading from the pharynx to the stomach), as in a belch. This speech has a characteristically 'burpy' quality, but it is used in a sophisticated way by many cancer-sufferers who have had a diseased larynx surgically removed. It is called *oesophageal* voice.

We can make other vocal effects, but they are better considered as emotional noises than speech sounds. For example, a short popping sound made with the lips, but with the sound sent outwards rather than sucked inwards, is fairly common in French, where, along with a distinctive hand gesture and shrug of the shoulders, it means roughly 'I couldn't care less' or 'It's not my fault'. A longer rasping sound, made by the tongue protruding slightly between the lips, is a signal of contempt in many languages – what in Britain is called a 'raspberry'. Some people flap the tongue noisily between their lips when they wish to show hesitation.

Finally, we should note that the vocal tract can produce several other kinds of sound which are not used in spoken language at all – or only in a highly idiosyncratic way. For instance, we can scrape or knock the teeth together, flap the tongue against the floor of the mouth, or make a sucking noise with the tongue against the inside of the cheek. If listeners hear such noises, they would not usually interpret them as attempts at communication.

The auditory-vocal channel of communication can evidently be put to work in the service of language in a variety of ways. But out of all the possibilities, one way stands supreme. Most of the sounds made by human beings in the 6,000 or so languages of the world (§51) use an outward flow of lung air. And the diversity of these sounds is made possible by a collaboration between larynx, mouth, and nose.

Two other cartilages work along with the thyroid to define the area of the larynx, and the movements of all three help to control the way the vocal folds vibrate.

The opening between the vocal folds, the *glottis*, is quite a small area. In men, the inner edge of the folds is usually between 17 and 24 mm; in women it is even smaller, from about 13 to 17 mm. But despite their small size, the vocal folds are remarkably versatile. Their tension, elasticity, height, width, length, and thickness can all be varied, owing to the complex interaction of the many sets of muscles controlling their movement. These movements take place very rapidly during speech, and produce several kinds of auditory effect.

How we use the larynx

To make voiced sounds

The most important effect is the production of audible vibration – a buzzing sound, known as *voice* or *phonation*. All vowels, and most of the consonants – such as [m], [b], and [z] – make use of this effect. It is in fact possible to feel the vibration. One way is to place the forefinger and thumb on either side of the Adam's apple, and compare the effect of saying a sound which is voiced, such as [zzz], with a sound which has no voice, such as [sss]. Alternatively, we can sense the resounding effect of vocal-fold vibration by making these sounds while putting a finger in each ear.

To make pitch movements

Each pulse of vibration represents a single opening and closing movement of the vocal folds. In adult male voices, this action is repeated on average about 120 times (or *cycles*) a second –

corresponding to a note on the piano about an octave below middle C. In women, the average is just less than an octave higher, about 220 cycles a second. The higher the pitch of the voice, the more vibrations there will be. A new-born baby's cry averages 400 vibrations a second.

An individual is able to alter the frequency of vocal-fold vibration at will, within certain limits, to produce variations in pitch and loudness which can convey contrasts of meaning. This linguistic use of pitch and loudness is described in such terms as *intonation, tone, stress*, and *rhythm*, and is discussed separately in §12.

To make glottal stops

The vocal folds may be held tightly closed – as happens when we are holding our breath. When they are opened, the released lung air causes the production of a *glottal stop*, heard very clearly in the sharp onset to a cough. A glottal stop is used as a consonant sound in many languages of the world. In British English, it is especially noticed in dialects that have been influenced by London speech (in such words as *bottle*, where it replaces the sound [t]).

To make glottal friction

If the vocal folds are kept wide apart, air expelled with energy will produce an audible hiss as it passes through the glottis – an effect that is often used as an [h] consonant sound in languages. Several paralinguistic vocal effects also use glottal friction, such as whispered and breathy voice (§3).

How we articulate

Once the air-stream passes through the larynx, it enters the long tubular structure known as the *vocal tract* (§4). Here it is affected by the action of several mobile vocal organs – in particular, by the tongue, soft palate, and lips – which work together to make a wide range of speech sounds. The production of different speech sounds through the use of these organs is known as *articulation*.

In addition, sounds produced within the larynx or vocal tract are influenced by the inherent properties of the cavities in the throat, mouth, and nose through which the air-stream passes. These cavities give sounds their *resonance*. Several kinds of resonance can be produced, because the vocal tract is able to adopt many different shapes. The most familiar one is nasal resonance, which is heard when we allow air to emerge through the nose. A 'nasal twang' is a feature of many regional accents, as well as of some physical disabilities.

When we describe articulation, it is usual to distinguish between those parts of the vocal tract that can move under the control of the speaker (the *active articulators*), and those which stay in one position at all times (the *passive articulators*). The chief passive articulators are:

- the upper teeth, especially the incisors, which are used to form a constriction for a few sounds, such as the first sound of *thin*;
- the ridge behind the upper teeth, known as the *alveolar ridge*, against which several speech sounds are made, such as [t] and [s]; and
- the bony arch behind the alveolar ridge, known as the *hard palate*, which is used in the articulation of a few sounds, such as the first sound of *you* [j].

All other organs are mobile, to a greater or lesser extent.

Articulating sounds with the pharynx

The *pharynx* is a long muscular tube leading from the laryngeal cavity to the back part of the oral and nasal cavities. It cannot be moved very much, but it is possible to make it narrower or wider, and this constriction can be used to make a consonant sound or to add a *pharyngeal* effect to another sound. *Pharyngealized* consonants and vowels can be heard in several languages, such as Arabic.

Articulating sounds with the soft palate

The *soft palate* is a broad band of muscular tissue in the rear upper region of the mouth. It is also known as the *velum* (the Latin for 'veil'). Its most noticeable feature is the *uvula* – an appendage that hangs down at the back of the mouth, easily visible with the aid of a mirror. In normal breathing, the soft palate is lowered, to permit air to pass easily through the nose – though of course the mouth may be open as well.

In speech, there are three main ways in which the soft palate affects the quality of sounds:

- it may be raised against the back wall of the pharynx so that air escapes only through the mouth; this produces a range of *oral* sounds – such as all the vowels and most of the consonants of English;
- it may be lowered to allow air to escape through the mouth and nose; this is the position required to produce *nasalized vowels*,as in French *bon* 'good', Portuguese, and many other languages; and

- it may be lowered, but the mouth remains closed; in this case all the air is released through the nose, resulting in such *nasal consonants* as [m] and [n].

Articulating sounds with the lips

The *lips* may be completely closed, as for [p] or [m], and this is the normal position for consonant sounds in English. But they may also be held apart in varying degrees to produce the friction of certain kinds of consonant, as in the *b* of Spanish *saber* 'know'. The lips also make the various kinds of rounding or spreading used on vowels – for example, the rounded lips of *boo* vs. the spread lips of *bee*.

Articulating sounds with the jaw

The mandible bone, which forms the *jaw*, permits a large degree of movement. It controls the size of the gap between the teeth, and strongly influences the position of the lips. Speakers sometimes adopt open or closed jaw positions – as when someone speaks 'through gritted teeth'.

Articulating sounds with the tongue

Of all the mobile organs, the *tongue* is the most versatile. It is capable of adopting more shapes and positions than any other vocal organ, and thus enters into the definition of a very large number of speech sounds: all the vowels, and most of the consonants. It is a three-dimensional muscle, the whole of which can move in any of three main directions:

- upwards/forwards, as in the [i:] of *bee* (the colon represents a long vowel);

also seen in the Romance languages, where all three notions are expressed by Latin *lingua* and its derivatives, such as French *langue*, Italian *lingua*, and Spanish *lengua*.

6
How we transmit sounds

When we speak, we produce energy in the form of sound. Sound energy is a pressure wave consisting of vibrations of molecules in an elastic medium – such as a gas, a liquid, or certain types of solid (e.g. along a telephone wire). For the study of speech production, the normal way of propagating sound is through the air. Air particles are disturbed through the movements and vibrations of the vocal organs, especially the vocal folds (§5). When we study speech reception (§7), air is not the only medium involved. The process of hearing requires the sound vibrations in air to be transformed into mechanical vibrations (through the bony mechanism of the middle ear), hydraulic changes (through the liquid within the inner ear), and electrical nerve impulses (along the auditory nerve to the brain).

When an object vibrates, it causes to-and-fro movements in the air particles that surround it. These particles affect adjacent particles, and the process continues as a chain reaction for as long as the energy lasts. If there is a great deal of energy in the original vibration, the sound that is produced may be transmitted a great distance, before it dies away. But the air particles themselves do not travel throughout this distance. The movement of each particle is purely local, each one affecting the next, in much the same way as a long series of closely positioned dominoes can be knocked over, once the first domino is pushed. However, unlike dominoes, air particles move back towards their original position once they have transmitted their movement to their neighbours. The movement is wave-like, backwards and forwards.

The way air particles move can be compared to a pendulum or a swing. At rest, a swing hangs down vertically. When it is put in motion, a backwards movement is followed by a forwards movement, on either side of the rest point, as long as there is energy available to keep the swing moving. This to-and-fro movement is called *oscillation*. Similarly, air particles oscillate around their rest point. As a particle moves forward, it compresses the adjacent particles and causes a tiny increase in the air pressure at that point. As it moves back, it decompresses these particles and causes a decrease in pressure. We can draw a graph of the pressure wave that is built up when particles move in this way. This graph is called a *waveform*. It is usual to draw waveforms as patterns from left to right, on either side of a horizontal line representing the passage of time. The simple movement of a single particle would look like this: Simple waveforms consist of a single pulse of vibration that repeats itself at a constant rate. The result is a *pure tone*. Pure tones are rarely heard in everyday life. Most sounds are complex, consisting of several simultaneous patterns of vibration. To produce a pure tone, you need a special electronic machine, or a device such as a tuning fork. When a tuning fork is struck, it vibrates with a single tone. The prongs of the fork move to and fro at a fixed rate. When the fork is held to the ear, we hear a pure tone.

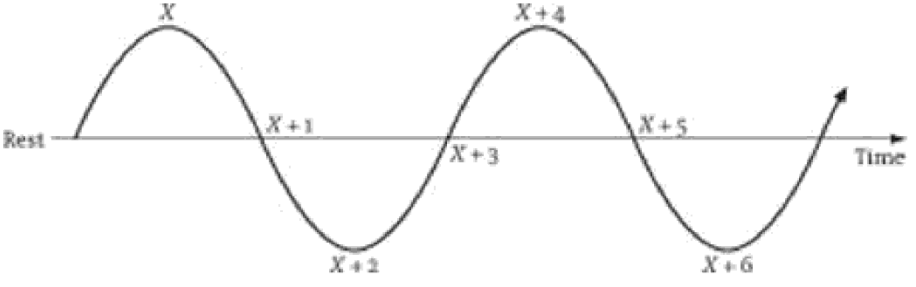

Fig. 3. Movement of a single air particle

Generating frequency

A single to-and-fro movement of an air particle is called a *cycle*, and the number of cycles that occur in a second is known as the *frequency* of a sound. Frequency used to be measured in *cycles per second* (*cps*), but this unit has been renamed *hertz* (abbreviated *Hz*), named after the German physicist Heinrich Rudolf Hertz (1857–94), who first broadcast and received radio waves. The basic frequency at which a sound vibrates is known as the *fundamental frequency*, generally abbreviated as *F0* and pronounced 'F nought'.

The range of frequencies that a young normal adult can hear is extremely wide – from about 20 to 20,000 Hz. It is not possible to hear vibrations lower than this (*infrasonic*) or higher than this (*ultrasonic*). However, the frequencies at both ends of this range are of little significance for speech: the most important speech frequencies lie between 100 and 4,000 Hz. The fundamental frequency of the adult male voice, for example, is around 120 Hz; the female voice, around 220 Hz (§5).

Frequency correlates to a large extent with our sensation of *pitch* – our sense that a sound is 'higher' or 'lower'. On the whole, the higher the frequency of a sound, the higher we perceive its pitch to be. But our perception of pitch is also affected by the duration and intensity of the sound stimulus. The notions of frequency and pitch are not identical: frequency is an objective, physical fact, whereas pitch is a subjective, psychological sensation.

One way of relating the physical notion of frequency to our sense of pitch is to relate familiar musical notes to fundamental frequency. Middle C has a frequency of 264 Hz. Middle A is the

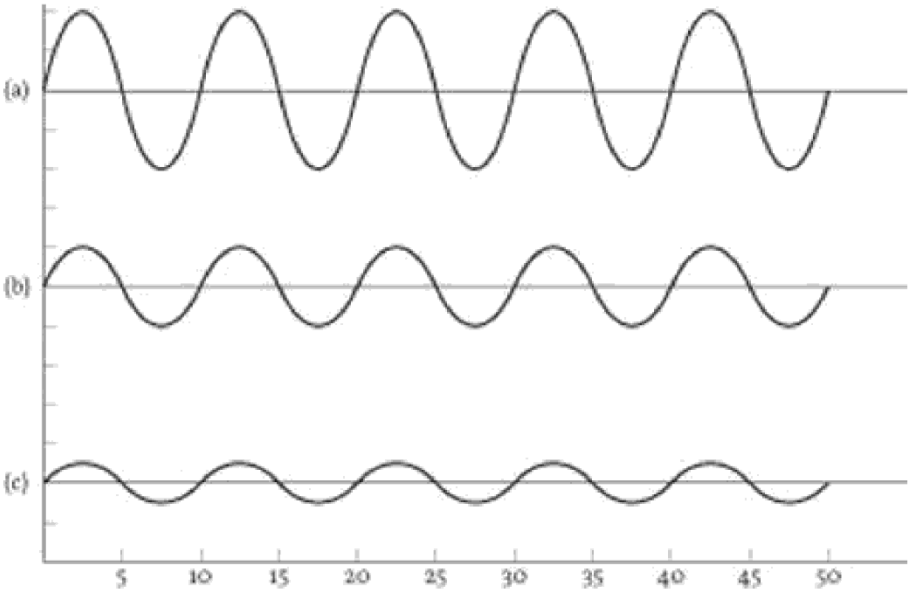

Fig. 4. Sine waves of equal frequency but different amplitude

We can get a sense of the decibel scale if we note the differences between familiar sounds.

- The rustle of leaves would be about 10 dB.
- The ticking of a watch held to the ear would be about 20 dB.
- A whispered conversation would be about 30 dB.
- An old-style typewriter in a quiet office would be about 40 dB.
- A car 10 metres away would be about 60 dB.
- Very busy city traffic would be about 70 dB.
- A noisy tube train would be about 80 dB.
- A pneumatic drill at 1 metre distance would be about 90

dB.

- An amplified rock band would be about 120 dB (at least).
- A four-engined jet aircraft at 30 metres distance would be about 130 dB.

At around 120 dB, the sensation of hearing is replaced by one of pain.

It is also possible to work out average intensity values for individual speech sounds. Vowels with the mouth wide open (such as [a:]) are the most intense sounds, followed by vowels made higher up in the mouth (such as [o:] and [i:]) and vowel-like sounds such as [r] and [l]. Much less intense are sounds involving a weak level of friction, such as [f], or those involving an articulatory closure and release, such as [p]. The decibel difference between adjacent sounds can be quite large. In a word like *thorn*, the increase in intensity from the first sound to the second is nearly 30 dB.

Making complex tones

Most sources of sound produce complex sets of vibrations, and this is always the case with speech. Speech involves the use of complex waveforms because it results from the simultaneous use of many sources of vibration in the vocal tract (§4–5). When two or more pure tones of different frequencies combine, the result is a *complex tone*.

There are two kinds of complex tone. In one type, the waveform repeats itself: a *periodic* pattern of vibration. In the other, there is no such repetition: the vibrations are random, or *aperiodic*. Speech makes use of both kinds. The vowel sounds, for example, display a periodic pattern; sounds such as [s] are

aperiodic.

It is possible to make an acoustic analysis of the complex wave involved in a particular sound and present its various components in the form of a sound *spectrum*. When we do this it becomes possible to see various 'peaks' of acoustic energy, reflecting the main points of resonance in the vocal tract. These peaks are known as *formants*, and they are numbered from lowest to highest: the *first* formant (F1), the *second* formant (F2), and so on. For a vowel like [i:], as in *bee*, spoken by a man at a fundamental frequency of 120 Hz, the F1 would peak at 360 Hz, the F2 at 2,280 Hz, and the F3 at 3,000 Hz.

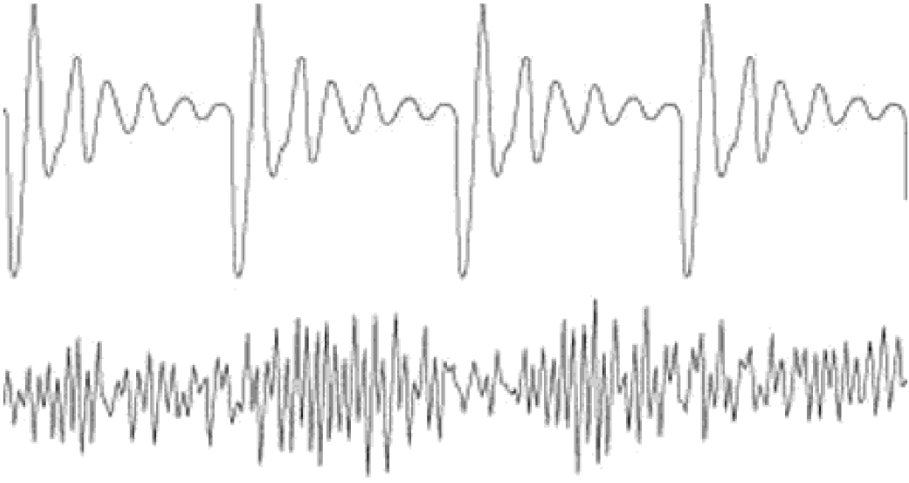

Fig. 5. Waveforms of a vowel [a:] and consonant [s:]

Formant structure is a major feature of speech sounds that involve vocal-fold vibration (§5) – which means all the vowels and all the voiced consonants, such as [b] and [n]. It is the formant pattern (especially the disposition of the first two formants) that enables us to tell vowels apart, or to recognize two vowels as being the 'same', even when produced by

different speakers. And vowel formants can also help in identifying the character of adjacent consonant sounds.

the bones of the middle ear, which is firmly attached to the membrane.

The chamber of the middle ear lies within the bones of the skull, about 15 mm high. It is filled with air, because there is a direct connection to the nose and throat via the *Eustachian tube* – named after the Italian anatomist, Bartolommeo Eustachio (1520–74). This tube is normally closed, but such activities as yawning or swallowing open it. In this way, the air pressure level on either side of the eardrum is maintained.

The primary function of the middle ear is to change the sound vibrations at the eardrum into mechanical movement – which will in turn be transmitted to the fluid-filled inner ear. It does this using a system of three tiny bones, known as the auditory *ossicles*. They are the smallest bones in the body, and the only ones to be fully formed at birth. They are suspended from the walls of the middle ear by ligaments, and are delicately hinged together so that vibrations can pass smoothly between them into the inner ear. The three bones have been given Latin names according to their shape: the *malleus* 'hammer', which is attached to the eardrum, the *incus* 'anvil', and the *stapes* (pronounced *stay-peez*) 'stirrup'. The stapes fits into the *oval window* – an opening in the bony wall separating the middle ear from the inner ear.

This may seem an unnecessarily complicated system of getting vibrations from point A to point B, but it is known to have several advantages. In particular, the process acts as a kind of leverage system, enabling the vibrations to be greatly amplified (by a factor of over 30 dB) by the time they reach the inner ear. As the inner ear is filled with fluid, vibrations would very readily get lost without this amplification. Also, the bony network of the middle ear helps to protect the inner ear

from sudden, very loud sounds. The muscles that control the movement of the eardrum and the stapes function in such a way that they lessen the chances of massive vibrations damaging the inner ear. However, the time it takes for these muscles to react is not so rapid that the inner ear can be protected from all such sounds, and cases of damage to the eardrum or inner ear do occur.

The inner ear

This is a system of small interconnecting cavities and passageways within the skull. It contains the *semi-circular canals*, which control our sense of balance, and the *cochlea*, a coiled cavity about 35 mm long, resembling a snail's shell. The main function of the cochlea is to turn the mechanical vibrations produced by the middle ear into electrical nerve impulses capable of being transmitted to the brain.

The cochlea is divided along most of its length into an upper chamber (the *scala vestibuli*) and a lower chamber (the *scala tympani*), separated by the *cochlear duct*. Both chambers are filled with a clear, viscous fluid known as *perilymph*. Vibrations enter this fluid via the oval window and the scala vestibuli, and are transmitted all the way around the cochlea. They pass from upper to lower chamber through an opening in the cochlear duct at its apex, and finish at a sealed opening in the wall of the middle ear, called the *round window*.

The cochlear duct is separated from the scala tympani by the *basilar membrane*, and is filled with fluid known as *endolymph*. This membrane is very thin at the base of the cochlea (about 0.04 mm) and gets thicker as it approaches the apex (about 0.5 mm). It is thus able to respond differentially to incoming vibratory pressures: high frequencies (§6) primarily

affect the narrow end; certain low frequencies activate the entire membrane.

Resting on this membrane is the highly sensitive organ of hearing, called the *organ of Corti*, discovered by the Italian anatomist Alfonso Corti (1822–76), and it is this which translates the mechanical movements of the membrane into nerve impulses. It contains a systematic arrangement of cells covered with very fine hairs, distributed in rows and layers along the membrane. These *hair cells* act as sensory receptors, picking up the pressure movements in the endolymph. Electrochemical changes take place, which activate the fibres of the auditory nerve. The signals are then sent the short distance along this nerve to the temporal lobe, via the brain stem and mid-brain, where they are interpreted as speech sounds.

Do we need two ears?

Two ears are a great asset. They enable us to be more precise in our judgement of the position of a sound source – an important factor in listening to people in a group, or heeding the direction of a vocal warning. This happens because a sound source is usually nearer one ear than the other; as a result, the signals to each ear will be slightly out of phase, and one will be more intense. The brain resolves these differences and makes a judgement about localization. Sometimes there is ambiguity (when a sound is reflected by a nearby object, for example), in which case we have to 'search' for the sound source by moving the head.

The value of two ears is most evident in cases of hearing loss in one ear. The 'good' ear copes well with a single speaker in a quiet room; but in contexts where sound is coming in from

several directions (such as in a meeting), the listener finds localizing the source of sound very difficult, and may look for the speaker in the wrong direction.

We should also note that the brain uses our two ears in different ways. One ear may have an advantage over the other for certain types of sound. This can be shown in tests of *dichotic* listening, where different signals are presented simultaneously to each ear, and listeners hear one sound ahead of the other. This shows that one ear transmits a sound to the brain more readily than the other. It is an important research technique in the study of speech perception.

waveform as a sequence of sounds and words. How is the brain able to analyse this signal so that the language units can be identified?

When we start to analyse the signal, we find other intriguing issues. If we hear different instances of a particular sound, we have no difficulty recognizing them as 'the same'. We hear the [b] sounds in bee, bay, bar, able, and *rob* as the same. But when we examine the relevant parts of the waveform, we find that a [b] before an [i:] vowel, as in *bee*, does not have exactly the same waveform as the [b] before an [A:] vowel, as in *bar*. Instances of [b] at the beginning of a word also differ from those at the end. Moreover, the articulation of [b] by different people will result in different waveforms because their regional accents and individual voice qualities will not be the same. It will vary, further, when people adopt different tones of voice (such as a whisper), or when it is said in a noisy situation. How does the brain recognize a [b] sound when there is so much variation?

In normal speech, people produce sounds very quickly (twelve or more segments per second), run sounds together, and leave sounds out. Nonetheless, the brain is able to process such rapid sequences, and cope with these modifications. For example, in the word *handbag*, the *nd* is actually pronounced as [m] in colloquial speech, because of the influence of the following [b]; but the word is still interpreted as *hand* and not *ham*. How does the brain carry out such partial identifications?

Looking for acoustic cues

One reason why we are able to recognize speech, despite all the acoustic variation in the signal, and even in very difficult listening conditions, is that the speech situation contains a

great deal of redundancy – more information than is strictly necessary to decode the message. The wide range of frequencies found in every speech signal presents us with far more information than we need in order to recognize what is being said. Just some of this information forms the relevant distinguishing features of the signal – features that have come to be known as *acoustic cues*.

The main research technique has been to create artificial sounds using a *speech synthesizer* – an electronic device that generates sound waves with any required combination of frequency, intensity, and time (§6). In the classic experiments using this device, the synthesizer was fed simplified acoustic patterns – a sound with two formants at certain frequencies, for example – and the researchers could then see whether the sound that emerged was recognizable as a certain vowel. Or, a sequence of formants, formant transitions, and bursts of noise could be synthesized, to see if listeners would perceive a particular sequence of consonant and vowel.

Using this technique, researchers found it was possible to establish the crucial role of the first two formants (§6) for the recognition of vowels. Similarly, the technique showed how we distinguish between voiceless and voiced consonants, such as [b] vs. [p] – it largely depends on the onset time of the vocal-fold vibration. And, in an important series of experiments, it was shown how the transitions of the second formant between a consonant and a vowel are especially important as a means of telling us where in the mouth the consonant is being made.

Such findings have laid the foundation for speech perception studies, and the way we perceive vowels and consonants is now quite well understood. But a great deal still remains to be explained. For example, the acoustic values cited

for the various sounds are averages, and do not take into account the many differences between speakers. Males, females, and children will produce the same vowel with very different formants, and it is not yet clear how we make allowances for these differences – for example, enabling us to judge that a male [a] vowel and a female [a] vowel are somehow the 'same'. Nor is it obvious how we handle the difference between stressed and unstressed sounds, or other modifications that result from the speed of connected speech.

How we perceive continuous speech

A great deal of research has been carried out on the auditory perception of isolated sounds, syllables, or words. In connected speech, however, very different processes seem to operate. We do not perceive whole sentences as a sequence of isolated sounds. And it turns out that the grammar of the sentence and the meaning of the words strongly influence our ability to identify linguistic units.

In one study, acoustically distorted words were presented to listeners both in isolation and in context. The context helped the listeners to identify the words much more accurately. In another study, single words were cut out of a tape recording of clear, intelligible, continuous speech. When these were played to listeners, there was great difficulty in making a correct identification. Normal speech proves to be so rapidly and informally articulated that in fact over half the words cannot be recognized in isolation – and yet we have little trouble following it, and can repeat whole sentences accurately.

Another feature of continuous speech perception is that we 'hear' sounds to be present, even if they are not. In one experiment, sentences were recorded with a sound

electronically removed, and replaced with a cough or buzz. Most listeners, when asked if there were any sounds missing, said no; and even if told that a substitution had been made, most were unable to locate it. In another study, people listened to one of four sentences, in which a sound (marked *) had been replaced by a cough, and were asked to identify a word which ended in *eel*.

It was found that the *eel was on the axle.

It was found that the *eel was on the shoe.

It was found that the *eel was on the orange.

It was found that the *eel was on the table.

People responded with *wheel*, *heel*, *peel*, and *meal* respectively, demonstrating the influence of grammar and meaning in perceptual decision-making.

Results of this kind suggest that speech perception is a highly active process, with people making good the inadequacies of what they hear arising out of external noise, omitted sounds, and so on. A further implication is that models of speech perception based on the study of isolated sounds and words are of little value in explaining the processes that operate in relation to connected speech.

Do we listen actively or passively?

There are two main views of speech perception. In one view, we are thought to play an active role in speech perception, in the sense that when we hear a message, we decode the sounds with reference to how we would pronounce them when we speak. Our knowledge of articulation (§5) acts as a bridge between the acoustic signal and the identification of linguistic

units. One major approach, proposed in the 1960s, is called the *motor theory* of speech perception. This theory argues that we model internally the articulatory movements of a speaker. We identify sounds by sensing the articulatory gestures that must have produced them – as if we were 'saying' words to ourselves to match the incoming speech. Another approach is known as *analysis by synthesis*. Here, we are assumed to make use of a set of mental rules to analyse an incoming acoustic signal into an abstract set of features. We then use the same rules to synthesize a matching version in production. Our perceptual system compares the acoustic features of the incoming signal with the ones it has generated itself, and makes an identification.

In the second view, listeners play a passive role. We simply hear a message, recognize the regular distinctive features of the waveform, and decode it. Listening is therefore essentially a sensory process, with the pattern of information in the acoustic stimulus directly triggering a response in the brain. No reference is made to a mediating process of speech production (except in difficult conditions, such as noisy speech situations). Several mechanisms have been proposed. One approach proposes a system of *template matching* – we match incoming auditory patterns to a set of abstract speech patterns (such as vowels and syllables) that have already been stored in the brain. Another suggests we use *feature detectors* – special neural receptors (analogous to those known to exist in visual processing) that are capable of responding to specific features of the sound stimulus, such as a particular formant, noise burst, or other general feature.

Both approaches have their strengths and weaknesses. Active approaches plausibly explain how we are able to adjust

9
How we describe speech sounds

The description and classification of speech sounds is the main aim of *phonetic science*, or *phonetics*. We can identify sounds with reference to their production (or *articulation*) in the vocal tract, their acoustic transmission, or their auditory reception. The most widely used descriptions are *articulatory*, because the vocal tract provides a convenient and well-understood reference point (§5); but auditory judgements play an important part in the identification of some sounds (of vowels, in particular).

An articulatory phonetic description generally makes reference to the following factors.

How we use the air-stream

The source and direction of air flow identifies the basic class of sound. The vast majority of speech sounds are produced using pulmonic egressive air (§4).

How we use the vocal folds

We need to consider the variable action of the vocal folds – in particular, the presence or absence of vibration (§5). *Voiced* sounds are produced when the vocal folds vibrate; *voiceless* sounds are produced when there is no vibration, the folds remaining open. Other vocal-fold actions are sometimes referred to, such as the way the glottis works when it produces a glottal stop (p. 27).

How we use the soft palate

We must note the position of the soft palate (§5). When it is lowered, air passes through the nose, and the sound is described as *nasal* or *nasalized*; when it is raised, air passes through the mouth, and the sound is *oral*.

Where we make the articulation

Place of articulation refers to the point in the vocal tract at which the primary closure or narrowing is made, such as at the lips, teeth, or hard palate. We may also need to take into account accompanying *secondary* constrictions or movements.

How we make the articulation

Manner of articulation refers to the type of constriction or movement that takes place at any place of articulation, such as a marked degree of narrowing, a closure with sudden release, or a closure with slow release.

How we use the lips

The position of the lips is an important feature of the description of certain sounds (especially vowels), such as whether they are *rounded* or *spread*, *closed* or *open*.

Whether we use other factors

In very precise descriptions of speech sounds, we may need to note other factors, such as the relative position of the jaw or the overall shape of the tongue.

Coarticulating

What is so impressive about speech is that all these factors are operating at the same time, and we describe a single speech

sound with reference to all of them. In addition, we have to remember that a 'single' speech sound is something of a fiction. The vocal organs do not move from sound to sound in a series of separate steps. Speech is a continuously varying process (§6), and sounds continually show the influence of their neighbours.

For example, if a nasal consonant such as [m] precedes an oral vowel such as [a], some of the nasality will carry forward, so that the onset of the vowel will have a somewhat nasal quality. The reason is simply that it takes time for the soft palate to move from the lowered position required for [m] to the raised position required for [a]. It is still in the process of moving after the articulation of [a] has begun. Similarly, if [a] were followed by [m], the soft palate would begin to lower during the articulation of the vowel, to be ready for the following nasal consonant.

When sounds involve overlapping or simultaneous articulations in this way, the process is known as *coarticulation*. If the sound becomes more like a following sound (its *target*), we are dealing with *anticipatory* coarticulation; if the sound displays the influence of the preceding sound, we are dealing with *perseverative* coarticulation. Anticipatory effects are far more common: a typical example in English is the way vowel lip position affects a preceding [s]. In such words as *see*, the [s] is pronounced with spread lips, anticipating the spread-lipped vowel. In such words as *sue*, the [s] is pronounced with rounded lips, anticipating the round-lipped vowel.

Using the International Phonetic Alphabet

The set of factors listed above is the basis of the International Phonetic Alphabet (or IPA). In 1886, a small group of language

teachers in France who had found the practice of phonetics useful in their work formed an association to popularize their methods. They called it the Phonetic Teachers' Association, and in 1897 the name was changed to the International Phonetic Association. One of the first activities of the Association was to develop the idea of a phonetic transcription, and the first version of the IPA was published in August 1888. The latest revision is reproduced on p. 54.

The chart shows all the consonants and vowels, as well as many diacritics used to identify subtle differences of pronunciation, along with symbols for clicks (§4) and pitch variations (§12). In the main table, we see the place of articulation of consonants changing as we move across the

Consonants (Pulmonic)

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

Where symbols appear in pairs, the one on the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Consonants (Non-Pulmonic)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	' as in:
| Dental	ɗ Dental/alveolar	pʼ Bilabial
! (Post)alveolar	ʄ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ʛ Uvular	sʼ Alveolar fricative

Suprasegmentals

ˈ Primary stress
ˌ Secondary stress — ˌfoʊnəˈtɪʃən
ː Long — eː
ˑ Half-long — eˑ
˘ Extra-short — ĕ
. Syllable break — ɹi.ækt
| Minor (foot) group
‖ Major (intonation) group
‿ Linking (absence of a break)

Tones and word accents

Level		Contour	
˩ or ˥	Extra high / Extra low	˩˥ or ˥˩	Rising / Falling
˦	High	˧˥	High rising
˧	Mid	˧˩	Low rising
˨	Low	˩˧	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

etc.

Vowels

Front Central Back

Diacritics Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

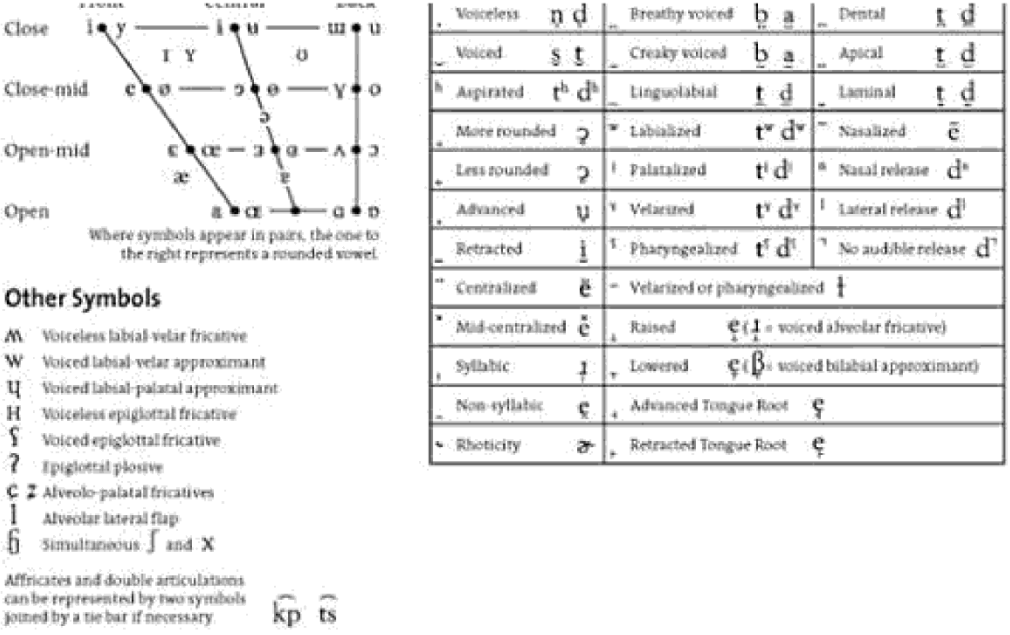

Fig. 7. The International Phonetic Alphabet

table from left to right, with the front of the mouth imagined to be on the left. The various manners of articulation are shown as rows in this table. Pairs of voiceless and voiced consonants are shown side by side, with the voiceless member on the left. Vowels, similarly, are shown in a diagram (§10) representing the central area of the mouth, with the vertical dimension showing tongue height and the horizontal dimension showing tongue position towards the front (on the left) and back (on the right) of the mouth.

The IPA looks complex, but this is because of its basic principle that there should be a separate letter for each distinctive sound, and that the same symbol should be used for that sound in any language in which it appears. The languages of the world display a very wide range of sounds, so most of the cells in the table are filled. On the other hand, there are


image
not
available

no such syllables as /mpf/ or /mfp/. Similarly, the sounds represented by *a*, *i*, and *e* are 'vowels' from both points of view: they are produced without audible friction; and they occur at the centres of syllables, in such words as *cap*/kap/, *hit*/hIt/, and *set*/set/.

However, there are a few problem cases. With the sounds usually written in English as *l*, *r*, *w*, and *y*, the two sets of criteria conflict:

- From a phonetic point of view, they are articulated without audible friction, and acoustically they display a similar energy pattern to that displayed by [a], [i], etc. A [w] is really a very short [u]. A [j] is a very short [i]. They must therefore be considered as vowels.
- From a linguistic point of view, these units typically occur at the margins of syllables, as in *let*/let/, *rat*/rat/, *wet*/wet/, and *you*/ju:/. They must therefore be considered as consonants.

The usual way of handling this problem is to say that these four units are neither consonants nor vowels but midway between these categories. They are, in short, vowel-like consonants, and might be described either as *semi-consonants* or *semi-vowels*. In practice, we usually describe [l] and [r] as *approximants* or *frictionless continuants*, and [w] and [j] as *semi-vowels*.

As an endnote to this section, I should emphasize that, in this part of the book, all talk of vowels and consonants is with reference to speech and not writing. In written English, for example, the 26 letters of the alphabet comprise 5 vowels and 21 consonants. In spoken English, there are 20 vowels and 24

consonants. It is this discrepancy, of course, which underlies the complexity of English spelling (§20).

10
How we describe consonants and vowels

Describing consonants

We usually describe consonants with reference to the four criteria described in §5.

- their place of articulation in the vocal tract;
- their manner of articulation in the vocal tract;
- the state of vibration of the vocal folds – whether vibrating (voiced) or not (voiceless); and
- the position of the soft palate – whether raised (oral) or lowered (nasal).

The present section concentrates on pulmonic egressive sounds (§4), which make up the vast majority of the sounds of speech.

Varying the place of articulation

Two reference points are involved in defining where we make consonants: the part of the vocal tract that moves (the *active* articulator) and the part towards which it moves or with which it makes contact (the *passive* articulator) (§5). Eleven possible places are used in speech.

Bilabial

We use both lips to make the articulation, e.g. [p], [b], [m].

Labiodental

We make the lower lip articulate with the upper teeth, e.g. [f], [v].

Dental

We make the tongue tip and rims articulate with the upper teeth, e.g. [T], [@], as in *thin* and *this* respectively.

Alveolar

We make the blade (and sometimes the tip) of the tongue articulate with the alveolar ridge (§4), e.g. [t], [s].

Postalveolar or *palato-alveolar*

We make the blade (and sometimes the tip) of the tongue articulate with the alveolar ridge, at the same time raising the front of the tongue towards the hard palate, e.g. [S], [Q], as in *shoe* and French *je* respectively.

Retroflex

We curl the tip of the tongue back to articulate with the area between the rear of the alveolar ridge and the front of the hard palate. Retroflex sounds are heard in many Indian English accents, and the 'dark' American English and British West Country use of *r* is often retroflex.

Palatal

We make the front of the tongue articulate with the hard palate, e.g. [ç], [j], as in German *ich* and *ja* respectively.

Fricatives

We make two vocal organs come so close together that the movement of air between them causes audible friction, as in [f], [z], and [h]. The term *fricative* reflects the nature of the sound produced.

Consonants which make an intermittent closure

Rolls or *Trills*

We make one articulator tap rapidly against another. Most often this is the tongue tip tapping against the alveolar ridge or the back of the tongue tapping against the uvula, and these articulations produce the different kinds of *trilled r*. Examples can be heard in the *r* sound of Welsh or Scottish English, as well as in many French and German accents.

Flaps

We make one articulator produce a single tap against another, as in some pronunciations of the *r* in *very*, or the *d* in *ladder*. In such cases the tongue tip taps once against the alveolar ridge. In Spanish a contrast is made between a trilled and a flapped *r*, as in *perro* [pero] 'dog' and *pero* [pero] 'but'.

Describing vowels

We usually describe vowels with reference to four criteria.

- the part of the tongue that is raised – front, centre, or back (§5);
- the extent to which the tongue rises in the direction of the palate. Normally, we recognize three or four

degrees: *high*, *mid* (often divided into *mid-high* and *mid-low*), and *low*. Alternatively, we can describe tongue height as *close*, *mid-close*, *mid-open* and *open*;

- the position of the soft palate – raised for oral vowels, and lowered for vowels which have been nasalized; and
- the kind of opening made at the lips – various degrees of lip rounding or spreading.

It is difficult to be precise about the exact articulatory positions of the tongue and palate because the tongue movements are very slight, and they give us very little internal sensation. We cannot easily feel where the tongue is in the mouth when we produce a vowel, though it is possible to develop this skill through phonetics training. Nor are absolute values possible (such as saying that the tongue has moved *n* millimetres in a certain direction), because mouth dimensions are not the same between speakers. We therefore tend to make vowel judgements on the basis of auditory criteria, in association with a limited amount of visual and tactile information.

The first widely used system for classifying vowels was devised by the British phonetician Daniel Jones (1881–1967). The *cardinal vowel diagram* is a set of standard reference points based on a combination of articulatory and auditory judgements. The front, centre, and back of the tongue are distinguished, as are four levels of tongue height:

- the highest position the tongue can achieve without producing audible friction;
- the lowest position the tongue can achieve; and

- two intermediate levels, dividing the intervening space into auditorily equidistant areas.

The grid provides a basis for vowel classification, along with information about the accompanying position of the lips. Jones called the main vowel-points 'cardinal' vowels, giving them numbers, and distinguished a primary series (1–8) from a secondary series (9–16), adding two further points (17–18). Each of these vowel-points was also given a phonetic symbol, and these are shown in the diagram (repeated here from p. 54).

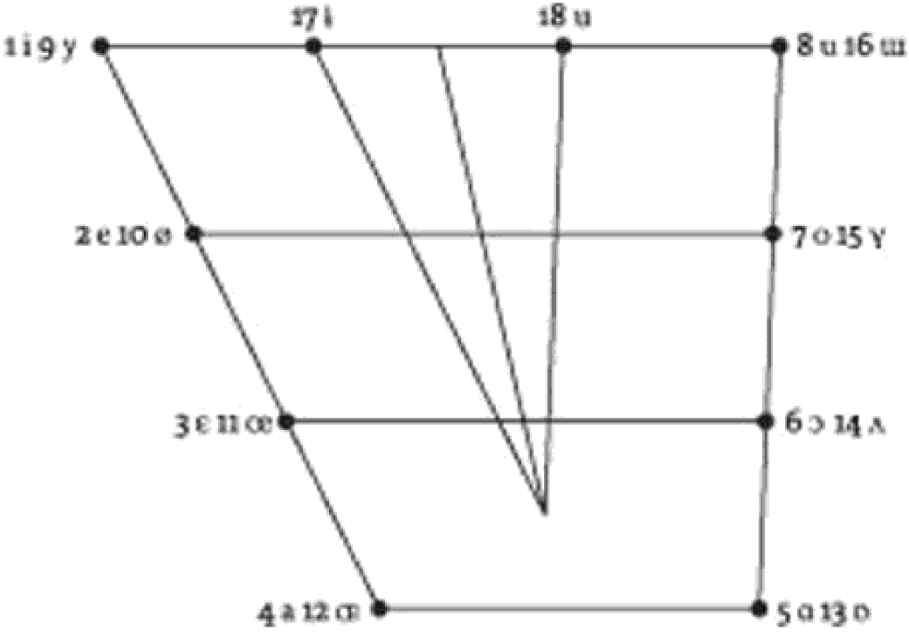

Fig. 8. The cardinal vowel diagram

The distinction between primary and secondary cardinal vowels is based on lip position. The first five primary vowels are all unrounded: front [i], [e], [Z], and [a], and back [A]. The remaining three back vowels are rounded: [C], [o], and

[u]. In the secondary series, the lip position is reversed: the first five are rounded: front [y], [ø], [Œ], and [Œ], and back [a]. The remaining three back vowels are unrounded: [v], [v], and [w]. The two other vowels represent the high points achieved by the centre of the tongue: they are unrounded [i] and rounded [u].

It should be emphasized that the cardinal vowels are not real vowels: they are invariable reference points (available on recordings) that have to be learned by rote. Once we have learned them, we can use them to locate the position of the vowels of any speaker. The above diagram shows the location of the eleven short and long vowels in the accent of English known as Received Pronunciation.

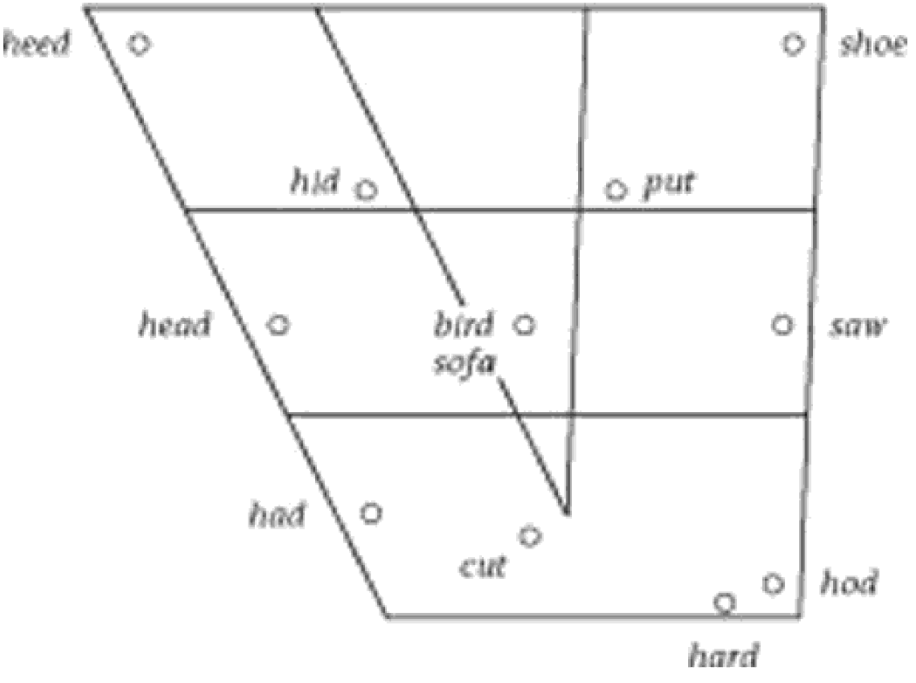

Fig. 9. Vowels in Received Pronunciation

Languages frequently make use of a distinction between

two kinds of vowels. In one kind the quality of the vowel remains constant throughout the articulation, and these are known as *pure vowels*, or *monophthongs*. In the other kind, there is an audible change of quality during the vowel, and these are known as *vowel glides*. If we make a single movement of the tongue during a glide, we call the effect a *diphthong*. If we make a double movement of the tongue, we call it a *triphthong*. Diphthongal glides in English can be heard in such words as *say* [seɪ], *fine* [faɪn], *cow* [kɑʊ], *boy* [bɔɪ], and *so* [səʊ]. Triphthongal glides are found in certain pronunciations of such words as *fire* [fɪə] and *power* [pɑʊə] – often represented in literary writing in such a way as *fiyuh* and *powuh*.

Describing slight differences

There are many tiny differences in the articulation of vowels and consonants. An [e] or [p] sound in one language may not be made in exactly the same way as the [e] or [p] sound in another. Phoneticians have therefore devised a set of symbols which can be used to show these very small differences. They usually take the form of a small *diacritical* mark, such as an accent, dot, or dash, attached to a more prominent symbol.

For example, when we make a consonant such as [s] or [d] we do not normally round the lips; but it is perfectly possible to produce the sound with a rounded quality. Indeed, this happens routinely in English when these consonants are followed by a rounded vowel, such as the [u:] in *soon* and *do*. If we want to draw attention to this 'secondary' feature of articulation, we can do so by attaching a diacritic to the main symbol, as in [s^{w}] or [Š].

Similarly, we can show the way we vary the articulation of a vowel by using diacritical marks:

11
How we organize the sounds of speech

Phonetics is the study of how speech sounds are made, transmitted, and received (§9). It is a subject that requires as its source of data a human being with a functioning set of vocal organs. The person's particular language background is not strictly relevant: phoneticians would draw the same conclusions about the production of speech whether they were studying speakers of English, Hindi, or Chinese.

But English, Hindi, and Chinese are very different languages, using sounds in very different ways. We therefore need a different focus when we study how languages use sounds, and this is what *phonology* provides. The aim of phonology is to discover the principles that govern the way sounds are organized in languages, and to explain the variations that occur. We begin by analysing an individual language to determine which sound units are used and which patterns they form – the language's *sound system*. We then compare the properties of different sound systems, and work out hypotheses about the rules underlying the use of sounds in particular groups of languages. Ultimately, phonologists want to make statements that apply to all languages.

The distinction between phonetics and phonology can be seen from a second point of view. The human vocal apparatus can produce a very wide range of sounds; but only a small number of these are used in a language as units to construct all of its words and sentences. Some languages use very small numbers of sound units – Rotokas, in the Pacific Islands, has only 11. By contrast, !Xũ in southern Africa has 141. English (in some accents) has 44. Whereas phonetics is the study of *all*