6

Language

- Man is the only animal who has a language that shows arbitrary symbols which can be combined in an infinite number of ways. In chimpanzees, evidence for true language is very controversial.
- The basic unit of speech is the PHONEME, of which there are different sets in different languages.
- Phoneme perception, which shows CATEGORICAL PERCEPTION, is probably carried out by special cells in the temporal cortex.
- Infants are born with mechanisms for identifying phonemes from all human languages, but these mechanisms disappear if not exposed to appropriate speech inputs.
- Prosodic features such as INTONATION CONTOUR are important both in conveying stress and emphasis, and in regulating the interactions in conversations.
- NON-VERBAL COMMUNICATION, indicated by the eyes or by other body movements, is important in indicating emotion and in regulating interaction.

Language is fundamental to human society and perhaps even to human intellect. The brain and mind solve almost overwhelming problems in understanding spoken and written language, problems which can probably only be fully appreciated by those trying to program computers to extract meaning from spoken speech, in which, at present, success is far distant.

If language is defined as communication by ARBITRARY SYMBOLS, from which potentially *infinite* different statements can be produced, then it is probably unique to man. The dance of bees is not a language because the symbols are non-arbitrary and limited in scope. Bird song does have arbitrary symbols, but their scope is limited. The only possible exception is in chimpanzees, in which recent experiments have suggested that true language might exist. Nineteenth century attempts to make apes talk failed because the larynx and mouth are not physically suited to producing human sounds, and as a result modern researchers have either used deaf-and-dumb sign language or plastic cut-out symbols. Thus the monkey Sarah was taught the meaning of several hundred plastic symbols, which could be combined in grammatically formed sentences of some subtlety, understanding statements equivalent to 'If Sarah take banana then Mary no give Sarah chocolate', 'What colour is banana?' and a meta-linguistic statement, asking about the language itself, 'What colour is symbol for banana?'. Whether this behaviour is truly language is still not entirely clear, some psychologists arguing that the results could be explained merely in terms of conditioning (see, the 'Clever Hans' effect, described in Chapter 3).

Language can be sub-divided into several divisions; SYNTAX, to do with grammar (why 'the dog bit the man' and 'the man was bitten by the dog' are different), SEMANTICS, to do with meanings of words, PHONETICS to do with sounds in spoken speech, and PROSODICS, concerned with intonation patterns (which means that the words 'How do you do?' can be spoken with perhaps a dozen or more inflections of meaning). In this chapter, we shall be concerned principally with phonetics and prosodics. Syntax, particularly, is a technically complex area and is best omitted in an introductory account.

Speech sounds begin in the larynx, which rapidly 'chops' the expired air to produce a sound rich in harmonics, which then resonates in the tuned cavities of the mouth, pharynx and nose, their resonant frequencies being changed by moving the tongue, lips and soft palate. A SPEECH SPECTROGRAM (e.g. Fig. 6.1) shows that speech contains a series of bands of energy centred at different frequencies or FORMANTS, the lowest of which is called the FUNDAMENTAL or F1, and higher ones F2, F3, etc. In Figure 6.1, it is not obvious where many speech sounds stop and others start; the clarity and separation of words in spoken speech are generated by the brain, and are not present in the acoustic input. Speech components that are essential for producing particular sounds can be discovered by means of a SPEECH SYNTHESIZER, which produces a particular formant pattern, and subjects then say whether they hear it as a particular speech sound. Vowels are the continuous, relatively steady patterns of formants whereas consonants are transitory sounds. A consonant cannot be sounded on its own but must be sounded before or after a vowel; if the separate sound components of a consonant are sounded on their own they do not produce speech, but instead produce click-like noises. The formants comprising consonants are not fixed in their transitions, but depend upon the vowel following them (e.g. the /d/ sound in /di/ and /da/ in Figure 6.2).

The basic unit of spoken speech is the PHONEME, of which there are 44 in spoken English, and 120 or more in the languages of the world, some being common to most languages, others being present in only a few. In English, the phonemes broadly correspond with the letters, but are more extensive, so that 'ŋ' refers to the 'ng' sound at the end of 'thing', and /ə/ (called 'schwa/) is the 'er' sound at the beginning of 'about'. Phonemes which are different in one language need not be different in another, so that, for instance, the 'p' sound in 'up'

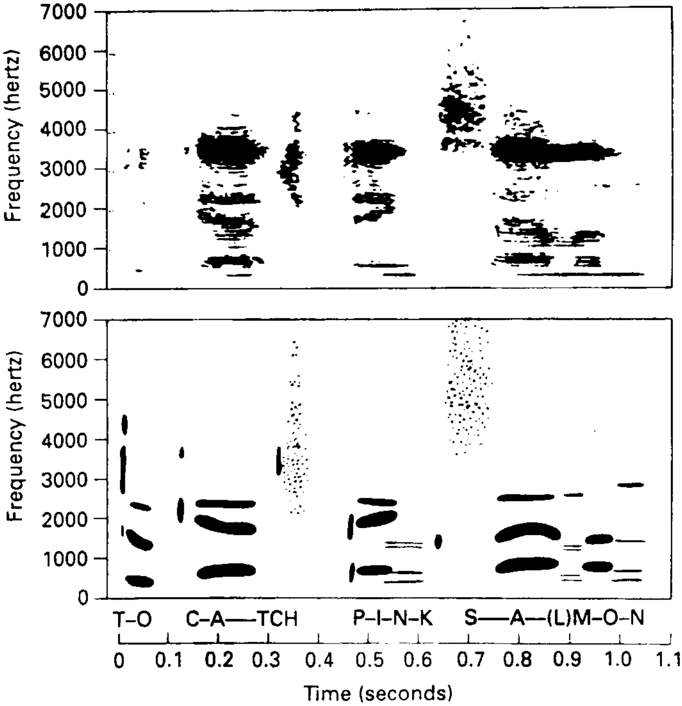

Fig. 6.1 An example of a speech spectrogram, in which time is shown on the abscissa, frequency on the ordinate, and amplitude is shown by the darkness of the record. The subject is saying the phrase 'To catch pink salmon'. The top box shows the actual spectrogram, and the lower box shows a diagrammatic representation, which can be put through the inverse process, using a SPEECH SYNTHESIZER or PATTERN PLAYBACK, to produce sounds which a normal listener will hear as the phrase 'To catch pink salmon'. Vowels (e.g. /ae/ in 'catch', or /i/ in 'pink') are seen to be extended continuous sounds made up of three formants. Consonants are transitory sounds preceding or following vowels, or are continuous noises (e.g. /s/ in 'salmon'). Reproduced from *Language and Speech*, p.66, by George A Miller. Copyright 1981 by W H Freeman and Company; reprinted with permission.

and 'lips' (technically aspirated and unaspirated respectively) are indistinguishable /p/ phonemes in English, but are separate phonemes in Hindi. The physical sounds making a phoneme may differ from occasion to occasion, but psychologically all are the same phoneme, in the same way that the letter A can be written differently by different people but is always the same letter. An analogy is with chemistry, phonemes being akin to atoms; chemistry depends only on which type of atom is involved, even though some of those atoms may have subtle physical differences in the form of isotopes. Thus $C^{12}O_2$ and $C^{14}O_2$ are chemically identical, although physically different. Just as

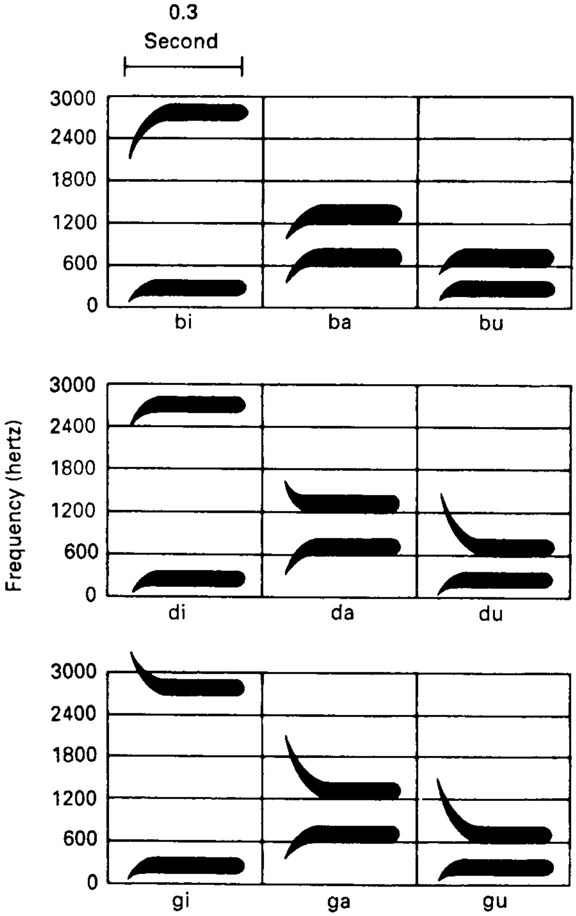

Fig. 6.2 Synthetic stimuli producing the consonants /b/, /d/ and /g/ before the vowels /i/, /a/ and /u/. The shape of the consonant depends upon the vowel following, e.g. the second (higher) formant of /d/ rises before /i/ and falls before /a/ and /u/. Reproduced from *Language and Speech*, p.66, by George A Miller. Copyright 1981 by W H Freeman and Company; reprinted with permission.

atoms are composed of smaller, sub-atomic particles, so phonemes are also composed of smaller units called FEATURES. These features, of which there are 13 in English, distinguish the phonemes. Thus, the stop consonants /b/, /d/, /g/, /p/, /t/, and /k/ differ in only two features: the PLACE OF ARTICULATION of the tongue, which is against the lips in the labial sounds, /b/ and /p/, against the teeth in the dental sounds /d/ and /t/, and against the soft palate in the velar sounds /g/ and /k/; and the onset of VOICING, the vibration of the larynx, which occurs after RELEASE of the consonant in the UNVOICED consonants /p/,

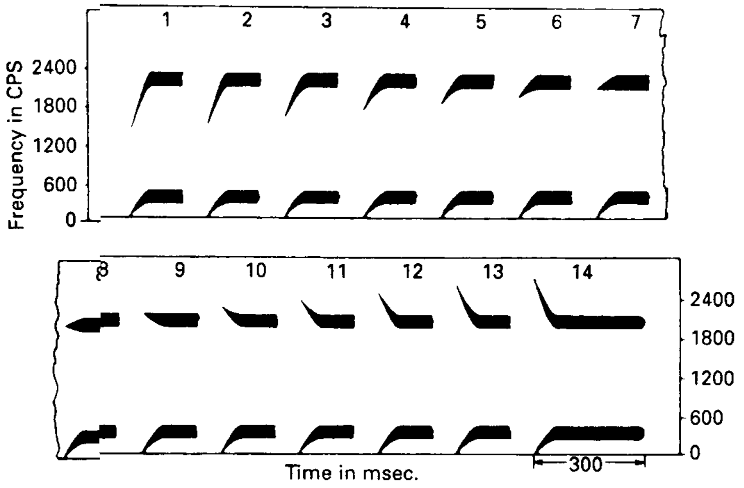

Fig. 6.3 *a*) A series of spectrographic patterns which at some points on the continuum are typical of naturally produced /bĕ/ (number 1), /dĕ/ (number 7), and /gĕ/ (number 14). From Lieberman A *et al.* (1957), The discrimination of speech sounds within and across phoneme boundaries, *J Exp Psychol*, **54**, 358–68.

/t/ and /k/, and before release in the VOICED consonants /b/, /d/ and /g/. Try speaking the sounds 'pat', 'tat', 'cat', 'bat', 'dat' and 'gat' and identifying where your tongue is placed and, by using your hand to feel the vibrations of your larynx, whether the sounds are voiced or unvoiced.

Perhaps the most exciting thing about phoneme perception is that it shows a fundamental difference from other forms of perception. If I play a sine-wave sound that gradually increases in frequency from 1300 Hz to 2900 Hz, you will hear a sound of gradually increasing pitch. By contrast, if I play the range of sounds shown in Figure 6.3a in which the initial frequency of the second formant shifts from 1300 to 2900 Hz you will not hear a continuously changing sound from /bĕ/ through intermediate values to /dĕ/ and then through more intermediate values to /gĕ/ but instead you will clearly hear each sound as *either* /bĕ/, *or* /dĕ/ *or* /gĕ/, with a very sudden changeover between the phoneme categories (shown in Fig. 6.3b). This CATEGORICAL PERCEPTION, in which all sounds are forced into pigeon-holes, without intermediate values, is almost entirely restricted to speech. There is good evidence that cells with such properties are located in the temporal lobe of the cortex, and that they can be ADAPTED by FATIGUE, in which repeatedly listening to the sound /bæ/ shifts the boundary between /bæ/ and /dæ/ (Fig. 6.3c). Damage to the temporal lobe can result in the deficit known as SPECIFIC WORD DEAFNESS or AUDITORY APHASIA in which a patient can hear and discriminate almost any sounds except speech sounds (see Chapter 23 for other types of aphasia).

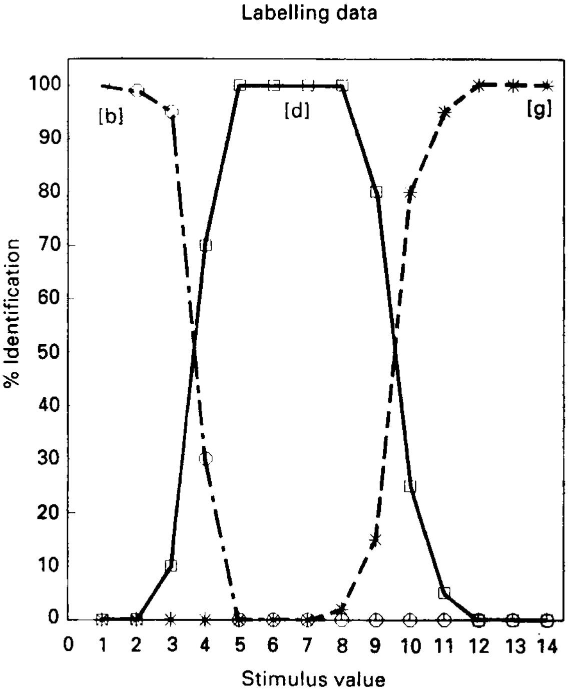

Fig. 6.3 *b*) Intermediate stimuli between /bĕ/, /dĕ/ and /gĕ/ are only actually heard as one of the three sounds, with no sense of blending between the possibilities. The ordinate shows the proportion of times that each sound in (a) was heard as /bĕ/, /dĕ/ or /gĕ/. Note the sudden and rapid changeovers. Adapted from Lieberman A *et al.* (1957). The discrimination of speech sounds within and across phoneme boundaries, *J Exp Psychol*, **54**, 358–68.

If phonemes have specific neural detectors, and languages differ in their phonemes, then how is it that an English speaker has the appropriate detectors for English phonemes? Two mechanisms might exist; listening to a language could induce appropriate detectors, or else *all* detectors may be present at birth and unused ones atrophy and disappear. The latter seems to be the case. By using either conditioning techniques or evoked potential methods, it can be shown that neonates can distinguish phonemes from birth onwards. Furthermore, they can also distinguish phonemes in languages other than their own, but the ability to make such distinctions disappears during the first year of life (Figures 6.4a and b). Such mechanisms may explain why second language learners can typically be distinguished from native speakers unless they have acquired their second language in the first years of life.

The correct use of spoken language requires not only that the words are enunciated correctly, but also that the overall INTONATION CONTOUR is appropriate. The fundamental pitch, or first formant, rises and falls continuously during speech, rising, for instance, at the end of a

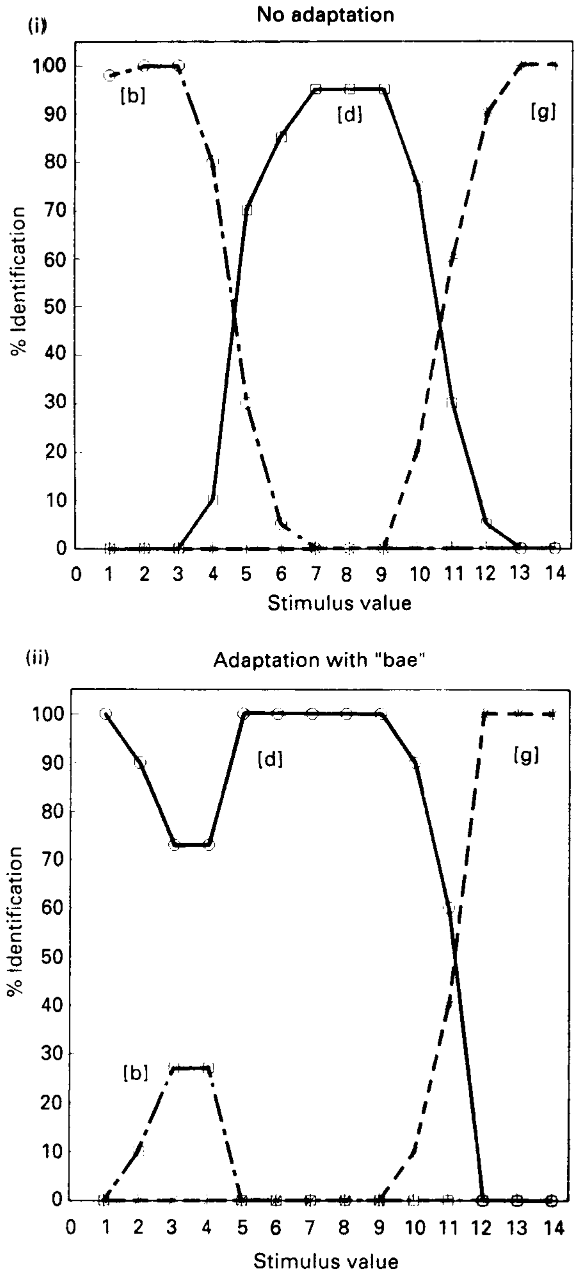

Fig. 6.3 *c*) The identification of a range of sounds between /b/ and /g/ by a single subject, *i*. before adaptation, and *ii*. after adaptation by listening to the sound /bæ/ 180 times during a 2 minute period, and then a further 1 minute of adaptation before testing on each stimulus. Note that the /d/-/g/ boundary is unchanged but that the /b/-/d/ boundary is disrupted, implying fatigue of a process detecting /b/ sounds. Adapted from Cooper W E and Blumstein S E (1974). A 'labial' feature analyzer in speech perception. *Perception and Psychophysics*, **15**, 591–600.

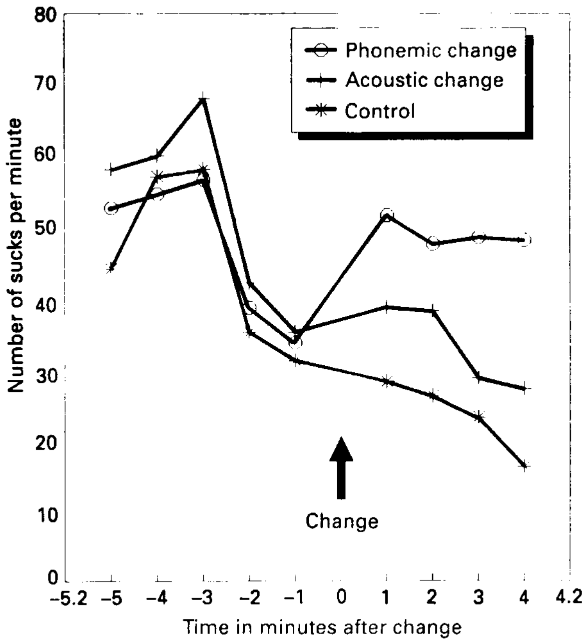

Fig. 6.4 *a*) The control condition shows the rate at which an infant sucked on an electronic dummy during a ten minute period. The number of sucks diminishes during the ten minutes (HABITUATION). A similar result applies in the 'acoustic change' condition (middle) in which one auditory stimulus is played repeatedly during the first five minute period and then an acoustically different (but not phonemically different) stimulus is played during the second five minutes. In the 'phonemic change' condition (left) the stimulus played during the second part of the experiment was a different phoneme from that in the first part, and habituation did not occur, indicating that the neonate must be categorizing the stimuli as being different, and hence must be perceiving the distinction between the phonemes. Adapted from Eimas P D (1985), The perception of speech in early infancy, *Scientific American*, **252**(1), 34–40.

questioning sentence and falling at the end of an emphatic sentence (and in writing we use '?', '!' and '.' to indicate such processes). These PROSODIC FEATURES tell another speaker when you are about to finish and hence when they should start, and they can also indicate when you will brook no interruptions. Like all such social conventions they can go wrong. A number of years ago it was noticed that Mrs Thatcher, who was then Prime Minister, was continually being interrupted by television interviewers. A careful analysis of the intonation contours (Fig. 6.5) revealed that at such 'disputed endings', Mrs Thatcher's voice pitch fell very quickly (normally a sign of an ending), and she looked directly at the interviewer (also a sign of an ending — see Fig. 6.6), but that her final pitch was not very

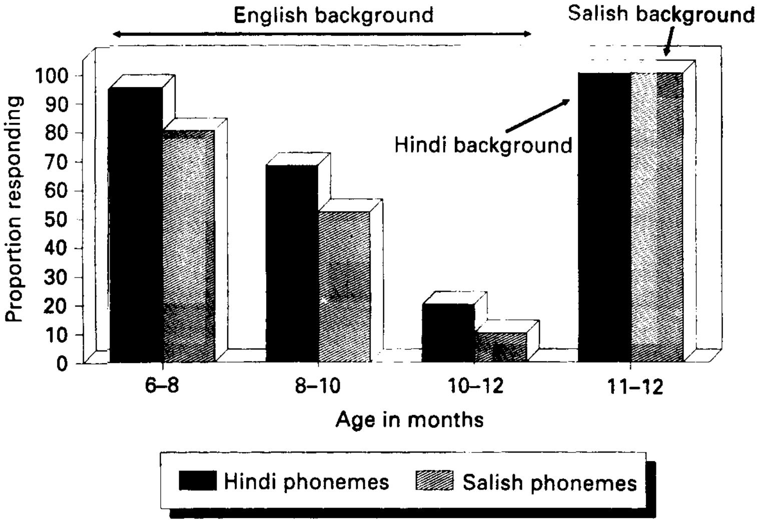

Fig. 6.4 *b)* Left three columns: The proportion of American infants at three different ages responding to a phonemic distinction that is either present only in Hindi (solid bars), or only in Salish (an American Indian tongue) (shaded bars). In native American infants the ability to perceive the distinction is present at 6 months but has disappeared by the age of 12 months. Rightmost column: the proportion of native Hindi or Salish infants who could make the Hindi or Salish distinction, respectively, remains at 100% at 12 months of age. Adapted from Eimas P D (1985). The perception of speech in early infancy, *Scientific American*, **252**(1), 34–40.

low (normally a sign of continuing to speak); the interviewers misinterpreted this combination of cues, thought that Mrs Thatcher must be finishing, and started to speak themselves, only to find that she in fact was continuing, having thought she had indicated that intention by keeping her final pitch high.

The use of direct eye-gaze to indicate the end of a speaking turn is only one of a whole set of NON-VERBAL COMMUNICATIONS presented by the eyes, which in most primates are surrounded by EYE-RINGS to make them more conspicuous — the eye-brows in man. So important are the eyes that when looking at faces we spend much of our time scanning them particularly. The eyes can convey direct information about emotional states (e.g. Fig. 6.7). They can also allow an inference about whether the truth is being told. Whilst deceiving, most people do not usually look the other person directly in the eye. An exception, in which direct eye-contact is maintained while lying, occurs in those individuals of the personality type known as HIGH MACHIAVELLIAN, after the Italian political theorist Niccoló Machiavelli (1469–1527), who refined deception as a political tool. The disconcerting effect of sunglasses upon interaction is in large part due to the abolition of eye-contact.

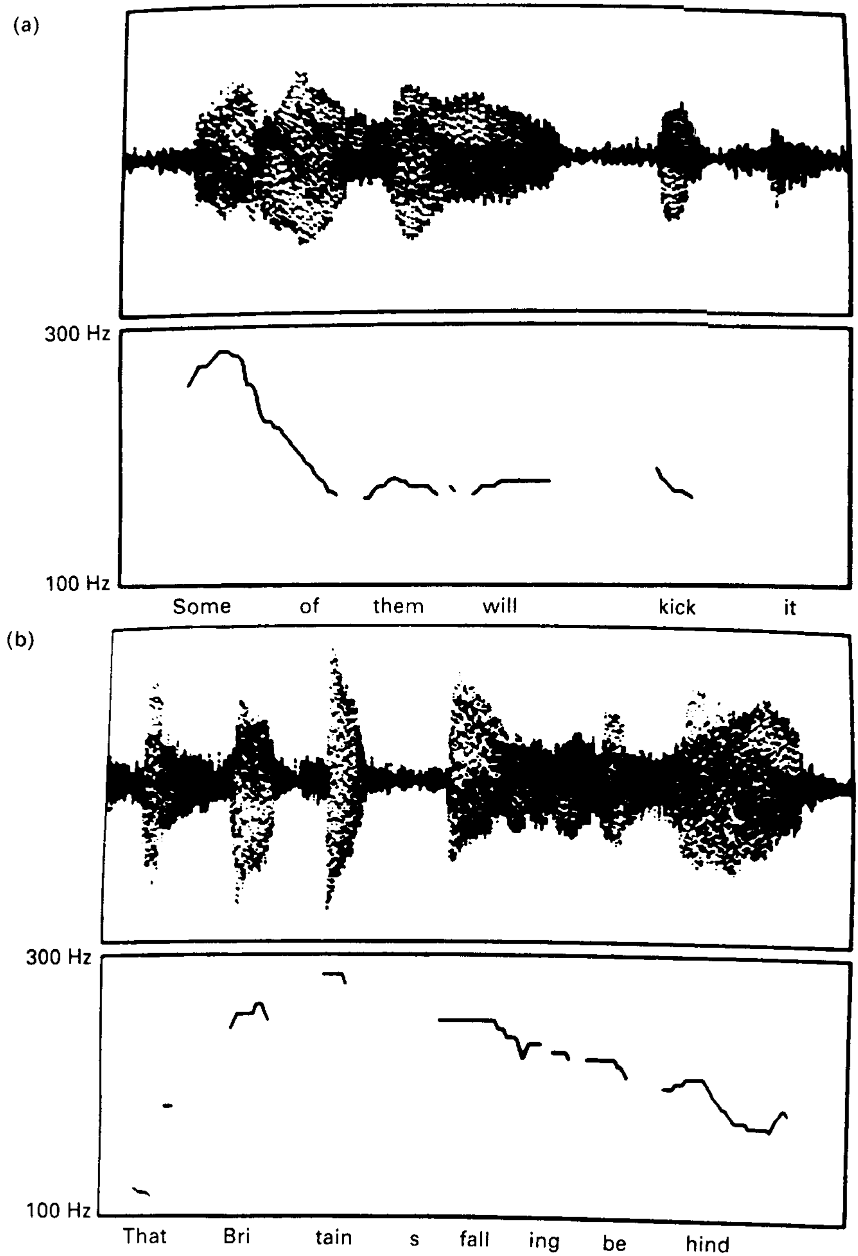

Fig. 6.5 The speech signals and the pitch of the fundamental frequency during two phrases spoken by Mrs Thatcher during television interviews. *a*) a FINAL UTTERANCE indicating the end of a bout of speaking, and for which the pitch falls quickly, and to a low level, and *b*) a MEDIAL UTTERANCE, with slower fall in pitch, and higher final pitch indicating an intention to continue speaking.

Reproduced with permission of authors and publishers from Beattie G W, Cutler A and Pearson M (1982), Why is Mrs Thatcher interrupted so often?, *Nature*, **300**, 744–7. Copyright 1982, MacMillan Magazines Limited.

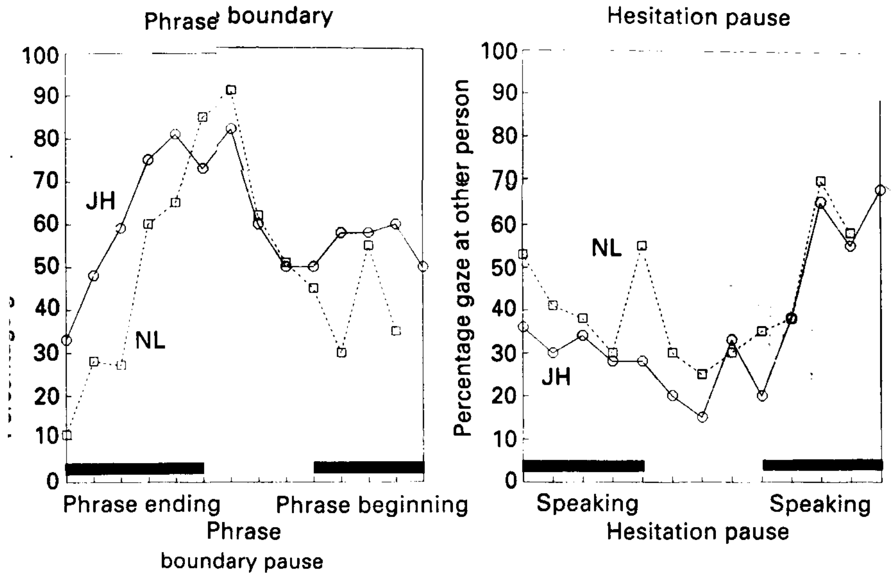

Fig. 6.6 The proportion of occasions on which two speakers are looking the other person directly in the eye during a total of 48 PHRASE BOUNDARIES (left), at which times a change of speaker can occur, and during 43 HESITATION PAUSES (right), in which the speaker intends to carry on speaking. Two different speakers, JH and NL, are indicated by separate lines. Redrawn with permission from Kendon A (1967), Some functions of gaze-direction in social interaction, *Acta Psychol Scand*, **26**, 22–63.

Fig. 6.7 The eyes are a direct source of information about emotional states. Reproduced with permission of author and publishers from Nummenmaa T (1964), *The language of the face*, Jyvaskala, University of Jyvaskala Studies in Education, Psychology and Social Research, Number 9.

In summary, spoken language is of crucial importance in human social communication, and as such special brain mechanisms have developed for its perception and production (see also Chapter 23), and these have been supplemented both by non-phonemic vocal mechanisms such as intonation, which conveys information about emotion and turn-taking, and by non-verbal mechanisms such as eye-gaze, for emphasizing linguistic features.