affection, bitterness, pleasure, disgust and so on. Each causes us to establish a different balance of forces in the larynx and hence to generate a different pulse wave, with different harmonic content and different sound quality. Similar differences, so extended in time as to constitute habitual modes of larynx action, characterize the voice and speech of individual speakers. When we refer to a speaker as a 'relaxed' or a 'tense' speaker, we are noting at least in part a difference in the extent to which physical tension is applied in the vibration of the vocal folds.

Summary of the larynx sound source

Vibration of the vocal folds, powered by air coming from the lungs during exhalation, is the sound source for voiced speech. It sets up a pulse wave in which the pulses are roughly triangular and of which the amplitude, fundamental frequency and waveform can be modified by the action of the laryngeal muscles. Mass, length and tension of the vocal folds are the physical factors which affect these variables; a speaker controls the balance between them in order to achieve the required effect at any moment. Fundamental frequencies in speech range from about 60 to 500 Hz but an individual speaker will not normally use more than about an octave. Men use the lowest fundamentals, women an intermediate range and children the highest. The switching on and off of vocal fold vibration is employed in the differentiating of voiced and voiceless sounds. The principal linguistic function of variation in fundamental frequency is the conveying of intonation patterns.

A triangular pulse wave has a spectrum in which all successive harmonics are present, with a progressive falling off of amplitude as frequency increases. Changes in fundamental frequency alter the spacing of harmonics in the spectrum but the general form of the amplitude spectrum remains approximately the same.

Apart from its role in conveying intonation and the voiced–voiceless distinction, the sound generated in the larynx does not transmit linguistic information. It acts as the carrier wave for this information which is imposed upon it by modifications introduced by the vocal tract; these modifications of the larynx wave are the subject of the following chapter.

7
The vocal tract

At the beginning of the previous chapter we saw that from the acoustic point of view the complete speech mechanism can be seen as a sound source coupled to a resonant system. The system is the air-way which leads from the larynx outwards through the pharynx and the mouth to the outer air, together with the path through the naso-pharynx and out through the nostrils when this branch is opened by the lowering of the soft palate. This is the system which is driven into forced vibrations by the pulse wave generated in the larynx. The sound waves radiated at the lips and the nostrils are the result of modifications imposed on the larynx wave by this resonating system. The first question therefore is: what are the properties of the vocal tract which determine these changes?

Forced vibrations, as we saw in Chapter 5, are the result of reflections and standing waves in the system, and these in turn are dependent on the natural frequencies and the damping of the system. It is these characteristics of the vocal tract that we need to examine.

Acoustic properties of the vocal tract

The example of the musical wind instruments showed that the dimensions of the air column involved were all-important in determining the frequencies at which resonance would occur. This must be so since the relation between the wavelength of sounds and these dimensions is the key to the phenomenon of resonance. The principle applies equally in the case of the vocal tract but the situation in speech is very much complicated by the fact that no two vocal tracts are the same size and shape and that short-term changes in the tract are brought about by articulatory movements. Communication by speech hinges on the fact that we learn to disregard the effects of the first and to pay close attention to the effects of the second type of variation.

In order to approach the problem of the acoustic performance of the

system with exactly the same amount of energy successively at different frequencies, the amplitude of the forced vibrations varies in accordance with the curve. Notice that in order to obtain this information we must be careful to drive the system with the same amount of energy, no matter what the frequency, otherwise we could not rely on the measured amplitude of the forced vibrations because some of the variations would be due to changes in the energy put into the system. The response curve specifies the acoustic behaviour of the system and remains valid for any driving force. In the case we are considering, the cylindrical tube of length 17 cm and cross-section 5 cm^2, with damping which approximates that of the human vocal tract, has three principal resonances at 500, 1500 and 2500 Hz. These will shape the forced vibrations or *output* of the system no matter what we drive it with. If the driving force is the larynx pulse wave, with the spectrum shown in Fig. 29, then the frequencies are all of different amplitude and the output will be the result of superimposing, as it were, the frequency response of the tract upon the spectrum of the glottal wave. In that case the output will have approximately the spectrum shown in Fig. 31(*a*).

The fundamental frequency of the larynx vibration in this example is 120 Hz and since all other frequencies in the complex tone must be multiples of this, there will not be a component at 500 Hz, which is the frequency of the first resonance of the tube. There is, however, a harmonic at 480 Hz which will show the greatest amplitude in the low part of the spectrum because of its proximity to the true resonance of 500 Hz. In the region of the second resonance of 1500 Hz, there is a component at 1440 and also at 1560 Hz. These are equidistant from the resonance but in the larynx spectrum 1560 Hz has lower amplitude than 1440 Hz and hence the latter will provide the second peak in the spectrum of the output. Near the third resonance there is a harmonic at 2520 Hz but there is also one of slightly greater amplitude at 2400 Hz; in the output spectrum we find two components of equal amplitude at this point, showing the presence of a true resonance between the two.

Let us now examine what happens when the fundamental frequency changes. Suppose that it rises to a value of 150 Hz. The components will now be the harmonic series 150, 300, 450, 600 Hz and so on. The lines in the spectrum will be more widely spaced and the peaks will appear at the places shown in Fig. 31(*b*), the first at 450 Hz, the second at 1500 Hz, since this is exactly the tenth harmonic of the fundamental, and the third at 2550 Hz. If the fundamental has a higher frequency still, let us say

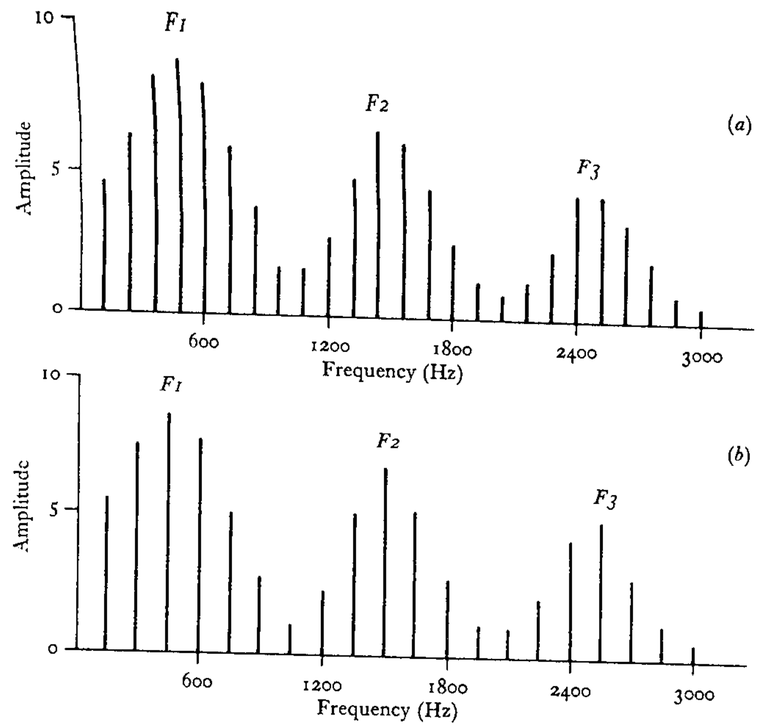

Fig. 31. Response of vocal tract to different fundamentals.

250 Hz, the spectral lines will be much wider apart still, but simple multiplication shows us that the resonances will fall exactly at 500, 1500 and 2500 Hz because there are harmonics of 250 Hz at each of these frequencies; they are the second, the sixth and the tenth harmonics.

In speech the fundamental frequency is changing all the time but the components of the larynx tone are always harmonics of the fundamental and the effect of the resonances of the tube or vocal tract is to produce a peak in the spectrum of the output at the harmonics which are the closest to the true resonance. This ensures that the spectrum of the resulting sound always has the same general outline or *envelope* although the fundamental frequency is continually changing. This fact is vitally important for speech because it means that a certain sameness of *quality* is heard in a range of sounds with different fundamentals. If this were not the case, speech sounds could not fulfil the linguistic function that they in fact have. The term used for a resonance of the system in this context is a *formant*. This is originally a German word used first by the physicist Hermann in the second half of the nineteenth century. The sound

vocal tract, we will begin with a drastic simplification of the conditions. The distance from the vocal cords to the lips in a male vocal tract is of the order of 17 cm. The area of any section across the tract varies greatly as we pass from the pharynx over the back of the tongue, under the hard palate and between the teeth but a representative area can be taken as 5 cm^2. Imagine the tract straightened out and formed into a cylinder of length 17 cm and cross-sectional area of 5 cm^2; since the section is circular, the diameter will be 2.5 cm. This cylindrical tube has the vocal cords at one end and may be regarded as being closed here, but it is open at the other end, that is at the lips. For such a tube, the first resonance will be the frequency of which the tube length is equal to the quarter-wavelength, that is a sound whose wavelength is 68 cm. The velocity of sound in air is 340 m/s, so that this frequency is 500 Hz. This means that if we try to drive the air in the tube with a range of different frequencies, beginning with some low frequency, the tube will absorb a large proportion of the energy fed to it until our driving force gets into the region of 500 Hz and here we shall get a peak in the amplitude of the forced vibrations. How sharp this peak will be depends, as we know, on the damping of the system and this we will consider later on.

The resonance at 500 Hz is not likely to be the only one, and if we extend the range of the driving force we shall find, for a tube open at one end and closed at the other, a second resonance at a frequency of which 17 cm is three-quarters of the wavelength and a third of which 17 cm is one-and-a-quarter times the wavelength. Since frequency is inversely proportional to wavelength, we can find these frequencies very simply by multiplying the first, 500 Hz, by 3 and then by 5, giving 1500 and 2500 Hz. The amplitude of forced vibrations falls off as frequency rises and we should probably not discover any appreciable resonance above this. The principal resonances due to reflections along the length of the tube, therefore, would be at 500, 1500 and 2500 Hz. The diameter of the tube, 2.5 cm, could only produce reflection of frequencies whose wavelength was comparable with this distance. Since these would be very high frequencies, we should not expect this dimension to influence the resonance characteristics of the tube.

In the previous discussion of vibrating air columns we paid no attention to the question of damping. This factor concerns the absorption of sound energy and clearly absorption is the opposite of reflection. The air will exhibit the same damping wherever we find it (apart from small fluctuations brought about by changes in temperature and humidity) but

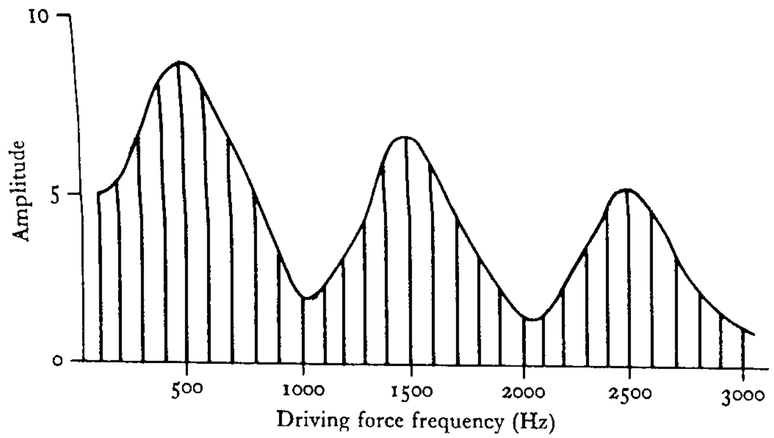

Fig. 30. Response curve of the vocal tract.

the material of the tube which encloses the air will have a great effect on the damping of the whole system. If the walls of the tube absorb a great deal of sound energy, the damping will be high and if they are efficient reflectors, there will be little damping. If therefore our cylindrical tube were made of glass or hard steel, the damping would be low, the system would be very selective or sharply tuned and the resonance peaks would be sharp. The material of the human vocal tract is of course rather the reverse of this; it is made of muscles and surface tissues, all of which are relatively absorbent and consequently the system will exhibit a high degree of damping. As we saw in Fig. 24, high damping results in a relatively flat resonance curve and the resonances of the vocal tract will be more of the form indicated in that figure than like the sharply tuned filter responses shown in Fig. 25. Because there is high damping, more energy will be absorbed generally and the overall amplitude levels will be reduced, but relatively there will be greater amplitude of forced vibration for the frequencies that intervene between the successive points of resonance. Consequently the three resonance curves will tend to join up with each other in the manner shown in Fig. 30. Each of the peaks indicates a resonance but the composite curve is most often referred to as a *frequency response curve* or *frequency characteristic* because it tells us how the particular system will treat a whole range of frequencies with which it may be driven. We must remember that the curve is obtained by joining up the highest points of lines giving the amplitude of spectrum components of the sound, in the way illustrated in the figure.

What the frequency response curve tells us is that when we drive the

section is longer, making *F2* much lower than for [i:]; in the spectrum shown in Fig. 32, *F1* is at 720 Hz and *F2* at 1200 Hz. The difference in configuration also produces some shift in the frequency of the third formant which is at 3000 Hz for [i:] and at 2520 Hz for [a:].

Formant structure is important because of the role that it plays in the recognition and differentiation of speech sounds. We have seen that changes of fundamental frequency produce a shift in the exact location of the peaks in the spectrum because these are tied to the harmonics but the formants, that is the true resonances of the vocal tract, will lead to spectral peaks in the same frequency region for a given configuration of the tract, regardless of changes of fundamental frequency. There are quite appreciable differences both in the range of fundamental frequencies and in the dimensions of the vocal tract as we go from one speaker to another, particularly as between men, women and children, but the general formant pattern enables listeners to recognize the 'same' vowel when it is uttered by many different speakers. The vowel of *heed* will always have *F1* and *F2* widely spaced and in the vowel of *hard* they will be close together.

It is possible by measuring enough spectra produced by a large sample of speakers to arrive at average values for formant frequencies. For vowel sounds generally, and this is true for the English system, a significant part of the information listeners use in distinguishing the sounds is carried by the disposition of *F1* and *F2*; Table 3 gives average values for *F1* and *F2* for the pure vowels of English based on data obtained from a sample of English speakers. From these figures it is possible to see something of the systematic relationship between formant frequency and articulatory configuration. The first four vowel sounds form a progression from a close front to an open front articulation. We saw from Fig. 32 that the first of these causes a wide spacing between *F1* and *F2*. As the articulation is made more open, the air passage over the hump of the tongue becomes wider and the constriction also moves towards the back of the mouth cavity. These two effects together produce a gradual change towards the equalization of the vertical and the horizontal sections of the tract with a consequent shift of the frequencies of *F1* and *F2* towards each other. When the point of articulation moves from front to back, as in going from the vowel of *had* to that of *hard*, there is a lowering of both formant frequencies. This apparently anomalous effect is explained by the fact that the tongue changes not only its position but also its shape; it can be bunched up so that the spaces both behind and in front of it become larger. The progression from open to close back vowel articulation

produces a gradual reduction in the frequency of *F1*; the sequence for *F2* is less regular as a result of the lip rounding which accompanies back vowels. The additional tube formed by the rounded lips lengthens the horizontal portion of the tract and lowers the second formant in comparison with an equivalent articulation with spread lips. This is particularly clear in the case of the vowel [o:] for which many English speakers use over-rounding of the lips; the mean frequency of *F2* is lower than for the neighbouring vowels. The two central vowels in the table, those of *hub* and *herb*, have formants which are intermediate between those of the front and the back vowels.

TABLE 3. *Mean frequencies of first and second formants of English vowels*

		F1 (Hz)	*F2* Hz)
i:	*heed*	300	2300
i	*hid*	360	2100
e	*head*	570	1970
a	*had*	750	1750
a:	*hard*	680	1100
o	*hod*	600	900
o:	*hoard*	450	740
u	*hood*	380	950
u:	*who*	300	940
ʌ	*hub*	720	1240
ə:	*herb*	580	1380

(After J. C. Wells, 'A study of formants of the pure vowels of British English', unpublished M.A. Thesis, University of London, 1962.)

The articulation of diphthongs involves a tongue movement from the disposition for one vowel towards that for another and these sounds therefore give rise to a more or less rapid switching from one set of formants to another. The sound in the word *how*, for example, will begin with the formant frequencies for [a:] and these will then change smoothly in the direction of the formants for [u:]; the sound in *here* will begin with the formants for [i] and change in the direction of those for [ə:] and so on for the other diphthongs. Just as the tongue movement in a diphthong is

produced by driving the 17 cm tube we have been discussing will have three formants, at 500, 1500 and 2500 Hz. The practice is to assign a number to a formant, beginning with the lowest one; in this example 500 Hz is the first formant, abbreviated as *F1*, 1500 Hz is the second, *F2*, and 2500 Hz is the third, *F3*. It should be noted that formants are strictly the resonant frequencies of the driven system but since a formant must give rise to a peak in the spectrum of the sound produced, the term formant is quite commonly applied to the frequency at which this peak occurs. Thus although *F1* of the 17 cm tube is at 500 Hz, in the spectrum of Fig. 31(*a*), the peak at 480 Hz might be labelled *F1*; similarly the peak at 1440 Hz may be labelled *F2* but *F3* must lie between 2400 and 2520 Hz because these two components have equal amplitude.

The cylindrical tube is not a very close approximation to the real vocal tract but if it actually had the damping of the latter, then the sound with the first three formants at 500, 1500 and 2500 Hz would sound remarkably like a central vowel of the [ə:] type. The notion of formants is particularly useful in connection with vowel sounds, though it can be applied to other types of sound. The arrangement of formants, what we may term the *formant structure*, is the basis for the recognition of most vowel differences. The peculiar property of the vocal tract is that its acoustic performance can be changed so as to bring about readily perceptible differences in formant structure. These changes are of course the result of differences in articulation which affect the shape and hence the dimensions of the vocal tract. The most important modifications are due to alterations in the configuration of the tongue. In the vocal tract itself the cylinder which has been used as an analogy is bent more or less into a right angle with a roughly vertical section in the pharynx and the back of the mouth and a roughly horizontal section in the front of the mouth. The body of the tongue can be moved backwards and forwards in such a way as to alter the relative length of the two sections and so change the resonances. The tongue also moves up and down and wherever it is highest in the mouth it forms a short tube which couples together the two sections of the vocal tract. The length and the cross-section of this tube have an influence on the formant structure of the sound and a further modification is introduced by the shape and the extent of the lip opening. Vowel systems in languages exploit all these means of creating differences in formant structure.

It must be stressed that the formant pattern of a particular sound is the outcome of the acoustic character of the whole tract working as one

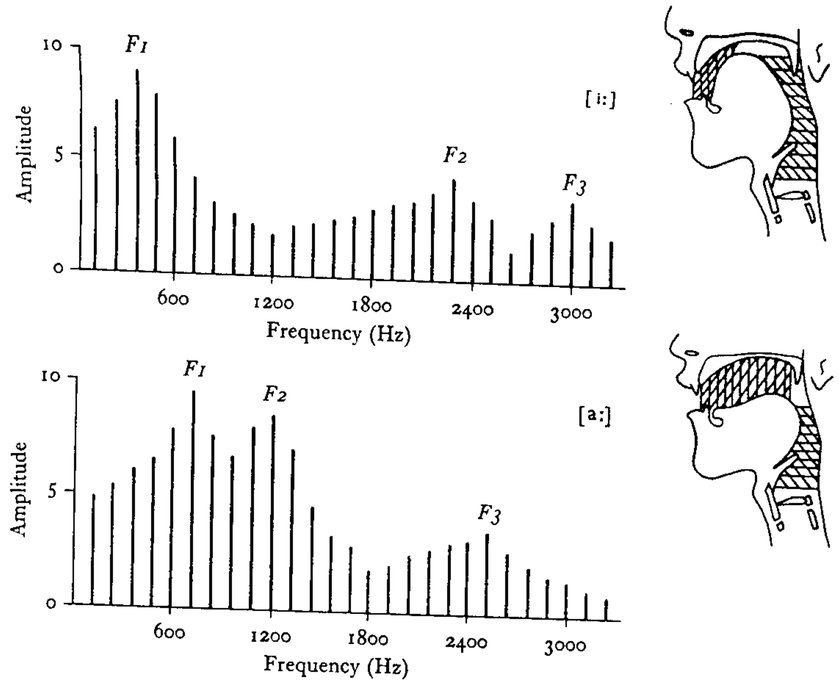

Fig. 32. Vocal tract shapes and spectra for [i:] and [a:].

resonant system. Hence it is not justifiable to assign any one formant to a particular part of the vocal tract. The frequencies of *F1* and *F2*, for example, are interdependent since in general the lengthening of one section of the tract implies the shortening of the other. It is true, however, that the vertical section is longer than the horizontal and is therefore responsible for the wavelength of the lowest formant, *F1*, while the shorter section tends to determine *F2*. The interdependence is well illustrated by the contrast between the two vowels [i:] and [a:]. Fig. 32 shows the shape of the vocal tract and also typical spectra for these two sounds. For [i:] the front of the tongue is high in the mouth so that the rear section of the tract is very long, while the part in front of the tongue constriction is very short. This results in an *F1* of comparatively low frequency together with an *F2* of high frequency. The spectra are shown for a fundamental frequency of 120 Hz, with *F1* at a frequency of 360 Hz and *F2* at 2280 Hz. A change to the articulation for the vowel [a:] entails a completely different configuration of the vocal tract. The narrowing of the air column is now towards the back of the mouth, so that the vertical section is shorter, thus raising the frequency of *F1*, and the horizontal

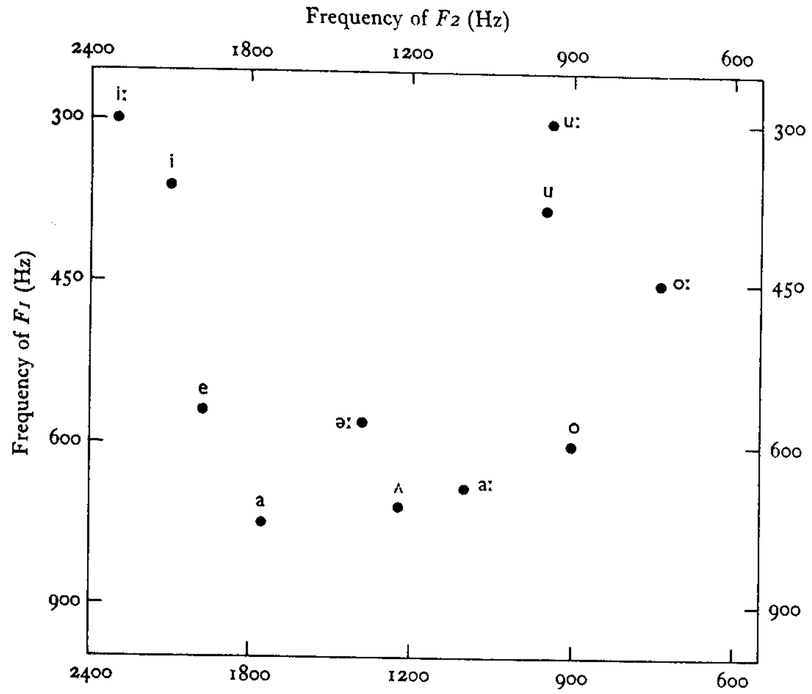

Fig. 33. *F1–F2* plots for English vowels.

not a complete change to a new vowel articulation but is rather in the nature of a glide in a given direction, so the modification of the formant frequencies may be of greater or less extent, depending upon the context and the speaker. It would therefore not be very meaningful to give mean formant frequencies for the elements in these sounds.

The relationship between articulation and formant structure has already been mentioned. This can be brought out rather strikingly by plotting a graph in which the frequencies of *F1* and *F2* for vowel sounds are represented on the two axes of the graph. The appearance of any graph is very much dependent upon the particular scales which are chosen for the related quantities; it may seem that those used in Fig. 33 are somewhat peculiar, but they are less so than they appear. The frequency scales for both *F1* and *F2* are logarithmic, that is to say that a multiplication of frequency is represented by the *addition* of a given distance on the scale. This serves to reflect more closely the impressions of relative pitch that a listener would gain; we saw earlier that the addition of equal pitch intervals requires the multiplication of the frequency of the

stimulus by a given factor: an octave jump in pitch means a doubling of the frequency, for example. The fact that the frequency of *F2* increases from right to left and that of *F1* from top to bottom of the graph does no more than determine the visual orientation of the pattern formed by the points that are plotted. Plotting the mean frequencies for the English vowels in this way in an *F1–F2* space immediately shows up a resemblance in general outline between the acoustic structure of the sounds and their articulatory character as it is reflected in the conventional vowel quadrilateral. Close front vowels, as we have already noted, have a low *F1* and a high *F2*; the distance between *F1* and *F2* decreases as we progress through the front vowels towards the open front articulation. The match between the vowel quadrilateral and the *F1–F2* plot cannot be exact because the former takes account only of the point of greatest tongue constriction while the formant structure is influenced also by the position of the whole tongue in the mouth and by the lip shape. These factors affect the back vowels in particular and here we see the greatest differences between the two representations; the *F1–F2* plot for [o:], for example, shows a low *F2* because of the lip-rounding.

It is therefore the acoustic characteristic of the whole vocal tract which modifies the wave generated at the larynx and hence shapes the sound which comes from the speaker's mouth. What makes the speech mechanism unique as a producer of sound is the fact that it is capable of an infinity of such modifications owing to the great flexibility of the articulators. We shall look at more of these possibilities in a later chapter when we come to examine in detail the acoustic characteristics of the various classes of sound in English. Before doing so, however, we must return to the subject of the generation of sound in speech, for we have so far considered only the vowel sounds, which normally depend upon larynx vibration as the sound source. There is a whole range of sounds, which include the voiceless consonants, which call for a quite different mode of operation.