consecutive harmonics occur over a wide frequency range and since the fundamental is low they appear close together. The fundamental of the clarinet note is an octave higher and therefore the distance between consecutive harmonics is doubled. It is only from about 1500 Hz upwards that consecutive harmonics appear in the clarinet tone; the second and fourth harmonics have zero amplitude. There are major differences in the mechanisms for generating sound in the piano and the clarinet: the piano tone is the result of free vibrations of the piano string which is struck by a hammer while the air column of the clarinet is performing forced vibrations in response to the continued vibration of the reed and does not show the rapid damping of the sound which is so characteristic of the piano. Nonetheless the differences in spectrum which appear in Fig. 26 are largely responsible for the difference in sound quality which we hear between the two instruments.

Differences in spectral pattern such as we see in this figure are usually recorded with the aid of a sound spectrograph, a device whose function is exactly what its name suggests, the drawing of a spectrum. The necessary acoustic analysis is done by means of electronic filters. The resonance curves of Fig. 25 could equally well represent the *response* of such filters tuned to 500, 1500 and 2500 Hz respectively. If a complex tone is fed into an analyser which includes these filters, then any component at 500 Hz or very close to it will be passed with very little loss by the filter and a measurement of the electrical output of the filter will give the relative amplitude of this component. Similarly the other two filters will register the amplitude of components in narrow bands centred on 1500 and 2500 Hz and of course the intervening frequency ranges can be explored in like manner to give a complete analysis of the sound. The patterns of Fig. 26 show one form that the resulting *spectrograms* may take and we shall see in later chapters that the spectrograph has proved an invaluable tool in the study of speech sounds.

6

The speech mechanism as sound generator

The acoustic principles formulated in the previous chapters provide a sufficient basis for a consideration of the acoustics of speech but, as we said earlier on, the sound waves of speech are among the most complex to be found in nature, particularly in the sense that extreme changes in sound quality follow each other with great rapidity. It follows from this that the speech mechanism viewed as a generator of sounds must work in a very complicated way and must indeed be capable of operating in a wide variety of ways. In this chapter we shall examine the basic properties of the speech mechanism in order to see what its principal modes of operation are and what kinds of sound they give rise to.

The musical instruments already used as examples consist essentially of a vibrating source of sound coupled to a resonating system and the speech mechanism is best considered in the same way. For a great deal of the time in speech the larynx is the *source* and the air column from the larynx to the lips, that is the *vocal tract* is the *system*. We know, however, that no sound can be produced without a supply of force or energy, so the first thing to establish is the nature of the energy supply for speech.

The energy supply in speech

Speech sounds may issue from a speaker in a continuous stream for quite an appreciable time and it is clear that the necessary force is not applied in one brief moment, as it is when a tuning fork is set in motion. It is the breathing mechanism, consisting of the lungs and the muscles of the chest and abdomen, that constitutes the energy supply. When we are not speaking, breathing is a rhythmical activity which involves taking air in and exhaling it on average about 15 times a minute, the actual rate depending upon the degree of physical activity. In quiet breathing the inspiratory and the expiratory phase of the cycle take about an equal time, so that if the cycle takes 4 seconds, 2 seconds is spent in

breathing in and 2 seconds in breathing out. In speaking we use only the expiratory phase to supply energy and we change the waveform of the activity so that breathing in takes a very short time and expiration is spun out to occupy a much longer time, which in extreme cases may be 10–15 s.

Air flowing out from the lungs constitutes the force used in generating speech sounds, as it does indeed in the case of wind instruments which are producing musical sounds. In both cases, in order that sound shall be generated, the steady flow of energy in one direction has to be converted into to-and-fro movements or oscillations, first of some solid structures and then of air particles. In woodwind instruments the mechanism is the reed or reeds which form the mouthpiece; in brass instruments the player's lips are the vibrating elements. For speech the larynx performs this function, with the vocal cords acting rather like the lips of the brass player.

The larynx as sound source

By the use of the laryngeal muscles the vocal cords can be brought together so as to form as it were a shelf across the airway which leads from the lungs through the trachea to the pharynx and the mouth. The effect is illustrated in Fig. 27(*a*). There is a steady flow of air from the lungs into the trachea and while the edges of the cords are held together, pressure on the under side of the shelf rises. When it reaches a certain level, it is sufficient to overcome the resistance offered by the obstruction and so the vocal cords open approximately in the way shown in Fig. 27(*b*). The ligaments and muscle fibres that make up these structures have a degree of elasticity and having been forced out of position, they tend to return as rapidly as possible to their initial disposition and thus to obstruct the flow of air once more. The pressure rises again and the cycle of opening and closing is repeated so long as comparable conditions in the larynx are maintained. The return of the vocal cords to the closed position is not due solely to their elasticity; it is aided by a phenomenon which is known as the Bernoulli effect, after the scientist who first described the principle: the rapid increase in airflow when the cords open results in a drop in pressure and a consequent suction effect which tends to draw the cords back into the closed position.

Since air is being continuously expelled from the lungs, the alternate opening and closing of the vocal cords results in the emission of successive puffs of air into the space above the larynx and it is this stream of pulses which is the basis of the sound generated by the larynx. The

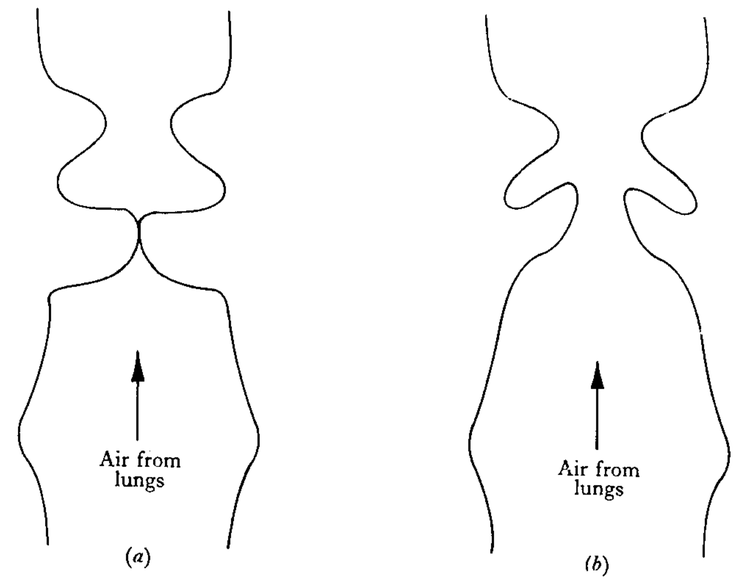

Fig. 27. Diagram of vocal fold motion.

time taken by one cycle of opening and closing of the cords depends upon the balance between the pressure below the cords, the *subglottal pressure*, and the resistance offered by the vocal folds, a convenient term for the ligaments and the muscles to which they are attached. This balance is under the control of the speaker and is the means by which we are able to change the frequency of vocal fold vibration at will.

Larynx frequency variation. The physical factors which regulate the frequency of vibration are the *mass*, *length* and *tension* of the vibrating structures, all of which are controlled by the intrinsic and extrinsic muscles of the larynx. The effect of these factors can be illustrated by the analogy of the stretched string, though it must be remembered that the vocal cords do not behave like strings. The strings which give out the low notes on a piano are obviously much longer and also much thicker and heavier than those which provide the top notes; it is clear that increasing the mass and increasing the length leads to a lower frequency of vibration. This is borne out by the orchestral stringed instruments, for the strings of the double-bass are longer and heavier than those of the violin. On the other hand, increasing the tension in a string raises the frequency of vibration; a string player sharpens the pitch of his string by screwing up a peg thus increasing the tension, and flattens the pitch by doing the opposite.

In the larynx mechanism, the vibrating parts can take on an infinity of different configurations and as a consequence the interaction between the effects of mass, length and tension is extremely complex. The closed phase of vocal fold vibration illustrated in Fig. 27(*a*) shows the folds in contact for a considerable proportion of their vertical height: they constitute a rather thick shelf. In other modes of vibration, the folds are thinned out at the edges so that the vibrating part is a great deal thinner and lighter; this reduction in mass leads to a higher frequency of vibration. The effective length of the vocal folds, that is the length which is actually set into vibratory motion, can be modified over a considerable range, chiefly by the action of the thyro-arytenoid muscles, which make up the body of the vocal folds, and of the crico-thyroid muscles, which change the angle between the thyroid and the cricoid cartilages and hence both lengthen and stretch the vocal cords. This last action provides a good example of the interaction of mass, length and tension in determining the frequency of vocal fold vibration. On the one hand it lengthens the cords and thus leads to a lowering of the frequency, but on the other hand it stretches the cords and this both increases the tension and thins the cords, both of which effects raise the frequency. For any 'setting' of the vocal folds, therefore, mass, length and tension are determined by muscle action and the fundamental mode of vibration is due to the resultant of the three factors.

The vocal folds have many modes of vibration in addition to the fundamental mode. Slow-motion pictures show for instance transverse waves travelling across the upper surface of the folds and when the contact between the two folds is extensive, as in Fig. 27(*a*), the opening of the folds begins at the bottom and travels upwards, so that there is a vertical wave-motion with a difference of phase between the opening of the lower and the upper surfaces. The wave generated by the whole larynx mechanism is therefore one of extreme complexity. The form of this wave has been arrived at by the use of several different techniques, principally by high-speed cinematography of the upper surface of the folds so as to register variations in the area between the edges of the two cords, and by measurement of variations in the airflow at the same place. The two methods give a wave which is substantially of the same general form, shown in Fig. 28. Time is represented on the horizontal axis and the vertical axis gives airflow in cubic centimetres per second. While the vocal cords are held in the closed position, no air is flowing at their upper surface and the corresponding part of the wave is a horizontal trace at the

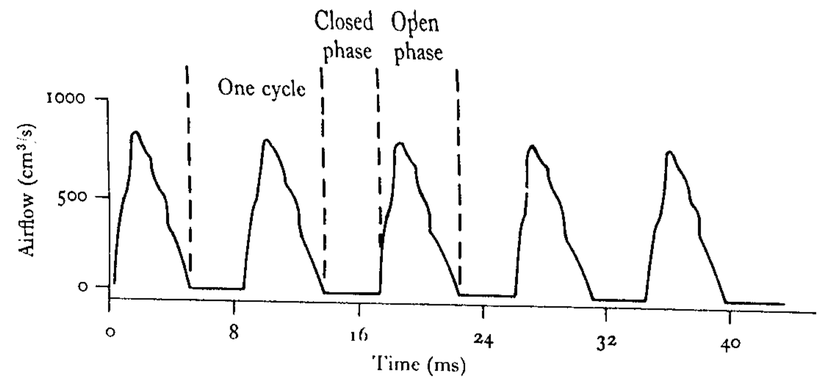

Fig. 28. Waveform of airflow at the larynx.

level of zero airflow. When the cords part, there is a rapid increase in airflow; the rise from zero to about 700 cm^3 takes just over 2 ms. As the cords begin to close together again, the airflow diminishes but at a somewhat slower rate, taking over 3 ms to return to zero, and it remains at zero for just over 3 ms before beginning the cycle again. The fundamental frequency is given by the time taken for one complete cycle; for the wave shown in Fig. 28 this is 8.3 ms, giving a frequency of 120 Hz.

The pitch of the larynx tone is determined by the frequency of the pulses emitted by the opening of the vocal cords; change in frequency means a change in the time occupied by one cycle, that is the closed phase plus the open phase. If the fundamental frequency were 100 Hz instead of 120 Hz, then the complete cycle of the glottal wave would take 10 ms; if it rose to 150 Hz, the time would be 6.7 ms and at 200 Hz it would be 5 ms.

The glottal wave is *periodic*, that is it consists of repeated cycles of motion, and therefore it must be made up of a fundamental frequency plus a range of harmonics. Consequently it must be possible to arrive at the waveform shown in Fig. 28 by adding sine waves together. The general shape of the wave is triangular and it may appear at first sight improbable that combining the curves we know as sine waves could ever result in such a shape. However, a sharp change in the direction of the wave implies simply that some event takes place in a very short time; given a sine wave of a high enough frequency, that is short enough wavelength, in the frequency mixture, sudden inflections can be produced in the complex wave. The quasi-triangular wave of Fig. 28 can be built up by adding an extended range of harmonic frequencies to the fundamental frequency so long as these harmonic components follow a

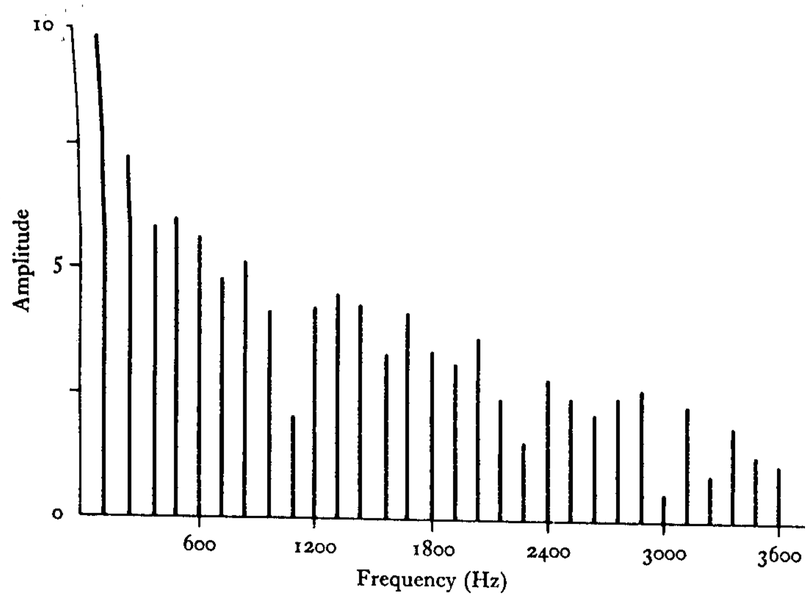

Fig. 29. Spectrum of the glottal tone.

certain general pattern of relative amplitude. The spectrum of Fig. 29 gives a good idea of what the amplitude pattern is for such a wave. Harmonics are shown up to a frequency of 3600 Hz, the 30th harmonic of 120 Hz, and the analysis has been stopped at this point; it does not mean that there is no energy at frequencies higher than this. The fundamental frequency has the greatest amplitude and each succeeding harmonic is represented, with the amplitude of harmonics falling off progressively as the frequency increases. This general trend is fairly regular but there are several dips in the spectrum indicating certain modes of vibration which have less amplitude than neighbouring ones.

The shape of the spectrum is not materially altered by a change of fundamental frequency but the spacing between the spectral lines is of course affected, as we saw in the case of the musical instruments. With a much higher pitch, such as might occur in a child's or woman's voice, the lines would be further apart and in a deep man's voice they might be closer together than those shown in the figure.

There are other features of the glottal wave which may vary, in addition to the fundamental frequency. The first is the overall or maximum amplitude of the wave. As we should expect, an increase in this amplitude will produce louder speech and a decrease, softer speech. The change is brought about mainly by increasing the general level of subglottal

pressure through the action of the breathing muscles. A second feature is the ratio of the open phase to the closed phase in the cycle. In the wave shown in Fig. 28, the time occupied by the open phase is 0.6 of the total time of the cycle. The value for this factor varies with conditions and with individuals. With higher fundamental frequencies the open phase takes up a great proportion of the cycle; this is understandable since the period of the cycle becomes shorter. In individual speakers who speak in a breathy way, the closed phase may be very short or there may indeed be no complete closure during the cycle; the narrowing down of the space between the vocal cords is enough to generate the sound of voice, though the mechanism is not in this case working very efficiently. The most efficient method of phonation is to produce the maximum amplitude of wave for a given subglottal pressure and this is brought about by reducing the time of the open phase compared with the closed phase. This not only increases the overall amplitude of the wave but also increases the relative amplitude of the higher harmonics in the spectrum. What was said earlier about the role of high frequencies in forming a triangular wave is relevant here. A shorter open phase means that the triangle has a shorter base and a more acute angle at the apex, and the sides are nearer to the vertical. All this means that changes in the vibratory motion take place in a shorter time, that is generate higher frequencies.

Functions of larynx vibration in speech

Since the primary object of speech is communication between people, the first requirement of speech sounds is that they should be audible. The basic function of the sound source in the larynx is to make speech audible. The glottal wave whose characteristics we have been considering is as it were the carrier wave of speech, that is to say it does not itself contain a great deal of information but it acts as the vehicle by means of which information can be conveyed, just as the radio wave picked up by a radio receiver is the carrier for speech and music and many other kinds of sound. From this point of view larynx vibration is important because it provides us with the means of ensuring, in ordinary circumstances, that our speech is audible to listeners. We are nearly always surrounded by noise of some kind and we use our own ears in controlling the amplitude of our larynx vibrations so that our speech shall compete adequately with the noise.

Fundamental frequency. The fundamental frequency of vocal fold

vibration is one of the most important aspects of larynx activity as far as speech and language are concerned. It varies continuously during speech and consequently the pitch of the voice never remains the same for any appreciable time. This is the essential difference between singing and speaking, for in singing the pitch of the voice is held steady for the length of one note, whether long or short and changes of pitch are made between notes. The range of fundamental frequencies used in speech depends upon the individual speaker, with men using the lowest range, children the highest and women an intermediate range. These differences are reflected in the average fundamental frequency for the three classes of speaker: 120 Hz for men, 225 Hz for women and about 265 Hz for children. The total range of fundamental frequencies encountered in speech extends from about 60 Hz to about 500 Hz. In an individual speaker the range ordinarily employed is not more than one octave and this is located in the lower part of his total voice range.

One mode of vocal fold vibration must be treated as a special case in considering fundamental frequency; this is the type of production known as 'creaky voice' by British and as 'vocal fry' by American investigators. It is heard at the end of utterances where fundamental frequency falls to a low level and is characterized by the interspersion of larynx cycles of abnormally long duration; sometimes such cycles alternate with cycles of shorter period, so that there is a short cycle followed by a long cycle, followed by a short cycle and so on. If the period of these long cycles is converted to frequency, they represent fundamental frequencies in the range from 20 to 60 Hz. The phenomenon is common but by no means universal among speakers and these frequencies have not been taken into account in the estimates of fundamental frequency given above.

The most important function of variation in fundamental frequency is as a carrier of intonation. Both the grammatical intonation patterns which form a part of every language system and the emotional variations imposed by individual speakers are conveyed very largely by this means. Like all linguistic elements, intonation patterns have to be recognized in spite of all the differences introduced by individual speakers and they therefore depend on relations of fundamental frequency and not on the use of any standard pitch scale. They are tunes which can be sung in any key and indeed with any scale and we each of us choose those that suit our particular make-up. The factor which has the greatest weight, therefore, is the *direction* in which fundamental frequency changes with time,

whether the pitch slides upwards or downwards and how rapidly it does so; after that the extent of the slide is important, but only in a rather gross way and in contrast with other patterns employed by the same speaker.

The circuits through which the brain controls speech activity enable us to adjust continuously the mass, length and tension of the vocal folds so as to produce the variations in larynx frequency corresponding to the intonation pattern we wish to convey.

Voice switching. A second linguistic function of larynx activity is connected with phonemic differences and contrasts. In running speech the vocal folds are vibrating for approximately 70 per cent of the time but owing to the occurrence of both voiced and voiceless consonants it is necessary for the vibratory activity to be switched on and off many times in the course of an utterance. The precise moment at which this switching takes place is very closely controlled by the speaker when he makes a distinction between pairs of voiced and voiceless sounds such as /p,b/, /t,d/, /k,g/, /s,z/ and so on.

Voice quality. The larynx waveform depicted in Fig. 28 is but one example of the kind of pulse wave that may be generated by vibration of the vocal folds. Since it has a particular arrangement of harmonics, that is a particular spectrum, it must constitute a particular sound quality. We can never actually hear the sound produced by a speaker's larynx because it does not become available to us until it has passed through his vocal tract and, as we shall see in the next chapter, been drastically modified by this transmission. It is, however, clear in principle that a change in larynx waveform must entail a change in quality. Individual speakers differ from each other in the formation and use of both the larynx mechanism and the vocal tract, but it has been established that differences in vocal fold vibration play a major role in enabling listeners to identify individual voices.

We have seen that within the speech of one individual there will be changes in the larynx wave with variation of fundamental frequency and of amplitude, particularly with regard to the ratio of the open to the closed phase of the cycle. The harmonic content of the wave will also vary according to the specific balance of mass, length and tension that is set up to produce a given frequency. The laryngeal muscles through which we adjust these factors are the medium by which we introduce into our speech the colouring expressive of our various moods and emotions,