SPEECH AND LANGUAGE

By D. B. FRY (London)

THE idea of the duality of speech and language is to be found everywhere in the literature that deals with communication between human beings. Sometimes it is implicit in what is said, at other times quite explicit, as it is for example in the writings of Hughlings Jackson and in the linguistic work of de Saussure, who labelled the dichotomy with the terms "la parole" and "la langue". The purpose of this paper is to give some account of the organization of speech and language in the individual, to show how they are connected and, at the same time, why it is valuable and indeed necessary to retain the distinction between the two terms.

The development of speech and language in the child is very closely bound up with the child's mental development. In the first year or two of life, the ordinary child develops the ability to perform the skilled movements which constitute speech and he learns to recognize and to interpret the acoustic signals which accompany these skilled movements when they are made by other people. During this period, he is also developing his mental powers in various ways, and among other things is moving in the direction of conceptual thinking. There is plenty of evidence that this development of mental capacity is very much dependent on the learning of speech. A child suffering from a congenital defect, such as deafness, which interferes with the learning of speech, does not develop the capacity to think in a normal way—or at least, he does not develop it at the normal rate. The classic case of Helen Keller, whose blindness and deafness prevented her from learning any means of communication in the normal way, offers some evidence on this point. She relates how at the age of 7 it first occurred to her that some common factor related the different forms of water with which she was familiar; in other words, she describes how the concept "water" first became a feature of her thinking, as the result of a step forward in her use of language.

The child's development of speech involves not only the learning of skilled movements but the gradual organizing at the psychological level of many processes that form the basis for both the transmission and the reception of speech. It is the events taking place at a psychological level that constitute the linguistic aspect of speech in the individual and which make it necessary to differentiate between speech and language.

Communication by speech has its origin in psychological activity in the speaker. There are, of course, different kinds of speech activity, but in the

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

common exchanges of speech between two individuals the purpose of the speaker is generally to influence in some way the mental activity of his listener so that it shall bear some particular relation to his own line of thought. The transferring of this influence is possible through the temporary establishment between speaker and listener of a communication chain (the speech chain) which is shown in Fig. 1.

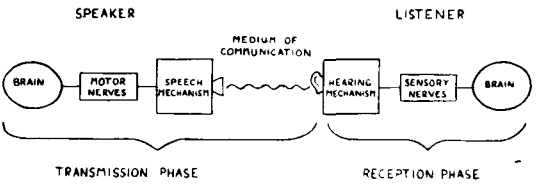


FIG. 1.
The speech chain.

Speech begins with activity in the speaker's brain which includes the impulse to speak, the emotional overtones that accompany the speech and above all the organization of what is to be said. This organization is carried out by the speech centres of the brain which then transmit the necessary nerve impulses through the motor nerves to the muscular systems involved in speech—the respiratory muscles, the laryngeal muscles, the muscles of the tongue, the pharynx, the soft palate, the lips. These muscles produce the skilled movements that constitute speech and, as a result of these movements, changes of pressure are set up in the air which serves as a medium of communication between the speaker and the listener. The pressure changes make up what we ordinarily refer to as sound waves and these are propagated through the air, arriving in the course of their travel at the ear-drum of the listener. Here the air pressure changes give rise first to mechanical movements of the ear-drum, the ossicles of the middle ear, the oval window, the fluid of the cochlea and the basilar membrane with its associated structures. The cochlear mechanism converts the mechanical movements into nerve impulses which are transmitted along sensory nerves to the brain of the listener where the speech centres complete the task of decoding the message that has been sent by the speaker.

The events that take place at different points in this chain present a great diversity—at one moment speech is a series of nerve impulses, at another, of muscle changes, or of air pressure variations. Speech consists, therefore, of a series of transformations—it is analogous to a telegraph system in which a message takes on many different forms in the course of its transmission and not to a postal system in which a letter passes unchanged from one stage to another until it reaches its destination. The events that take place in speech may be classified broadly as being of three kinds—those that take place at a psychological level, those that are mainly physiological, and those that are purely physical. In the transmission

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

phase of speech, psychological events are first transformed into physiological ones (through nerves and muscles) and the physiological are then transformed into physical events in the medium of communication. The reception phase consists in the unwinding, as it were, of the transformations effected during the transmission phase, the conversion of physical changes to physiological ones, of physiological to psychological. In successful communication by speech, there must clearly be features common to the events at opposite ends of the chain—that is, in the speaker's and the listener's brain. No matter what transformations take place in the course of a speech event, the listener must be able finally to convert the message into a form resembling that in which it was organized by the speaker's brain and he can do this only if he has in common with the speaker a great deal of previous experience—if both hold a large memory store of language material in common. When this is the case, the speaker and listener are said to speak the same language, and the material which is held in common, not only by individual speakers but by a whole community, makes up what de Saussure designated "la langue".

The Organization of Language

Every language is a system—a system of units of differing size which combine together in a manner regulated by the structure of the system. The smallest unit in spoken English is the *phoneme*, the equivalent in speech of the single letter in writing. The English word *hid* is made up of three of these phonemic units; the word *hit* is made up of three, of which two occur also in *hid*; the word *peg* is made up of three different ones again. The total number of English phonemes is just over 40, the exact number depending on the particular speaker. This number can be determined by making a series of substitutions in English words, like the change made to convert *hid* into *hit*. There is, for example, a series of English words of this type: *heed, hid, head, had* . . . in which the first and last phonemes are the same in each case, whereas the middle phoneme is different in each word. The middle phonemes, in this particular type of three-phoneme English word, are the vowel phonemes of the English language system. Not all English vowels can occur between /h/ and /d/, but by taking other series into account, we can decide the total number of vowels, which, for most English speakers, is about 20. The number of vowels, particularly, is dependent on the pronunciation of the speaker since, for example, an Englishman who thinks of different vowel phonemes in the words *gas* and *mask* will probably have one more in his total than the Englishman for whom they contain the same vowel phoneme. The English consonant phonemes can be arrived at by similar substitutions in initial and final position in words of the type: *pit, bit, kit, sit* . . . and this gives a list of 24 consonant phonemes.

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

The next linguistic unit in order of size is the *morpheme*. This term is used to refer to the units which have mainly a grammatical function in the language. For example, the verb *dictate* is a single morpheme; another morpheme *-ing* is added when we need the present participle of the verb, *dictating;* a different morpheme, *-id*, is added to make the past tense, *dictated*. If in the present tense we say *he dictates*, then the morpheme *-s* is added to make the proper form of the verb. Morphemes may, therefore, consist of a single phoneme or a sequence of phonemes. The complete inventory of English morphemes includes the root form of all words, with the addition of all particles that have a grammatical function.

The combining of morphemes produces the units that are generally recognized as words. Each form of the verb *dictate* just given is a single word, though most of them comprise two morphemes. It is a feature of all language systems that the procedure by which units of one magnitude are combined to form units of the next order is never entirely regular, and is purely a matter of function. /s/ is a phoneme in English, but there is a morpheme *-s* as well; there is a morpheme *dictate*, but there is a word *dictate*, too. The reappearance of superficially the same unit at different levels and with different functions occurs in all linguistic systems.

The largest unit in spoken English is the sentence and the process of combining words into sentences is too familiar to need any comment. It must be stressed, however, that in spoken English many speech sequences form sentences although they have not the complete grammatical form of a written sentence. The following exchange might take place in conversation: *Is it important? Not very*. The two words, *not very*, form a sentence in spoken English, since they constitute a complete remark. A sentence may equally well consist of a single word—the answer might have been the word *quite*. Many words of this type, like *yes*, *no*, etc., make single word sentences.

The knowledge of all these units and of the way in which they combine is stored in the memory of every speaker and listener, though, of course, we are not usually aware of this fact. We know not only how the units combine, but even a great deal about the statistics of their combination. We shall see later on how all this linguistic knowledge is used in the reception of speech.

The Organization of Speech

The function of speech is to convert language into a transmissible form and the reception of speech consists in recovering language units from the transmitted signals. In the processes represented in the speech chain, speech appears as a predominantly motor activity which has acoustic consequences, and it has its own organization which is described most appropriately in terms of physiological and physical events.

During speech, the speaker's brain is, as it were, feeding the speech

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

mechanism with a more or less continuous flow of language material which is to be converted into transmissible signals. The first effect of this process is a modification of respiration, since the motor activity of speech takes place only during the expiratory phase. In speech the rhythm of inspiration and expiration established in quiet breathing is broken. The expiratory phase is very much lengthened and inspiration is effected very rapidly at convenient moments. There are a number of factors which determine the moments convenient for inspiration; first, the linguistic factor: the pause for breath in normal speech coincides with the end of a sentence or sense-group; the speaker tends to keep up a continuous flow of speech until he reaches a point in the linguistic organization where a break is possible.

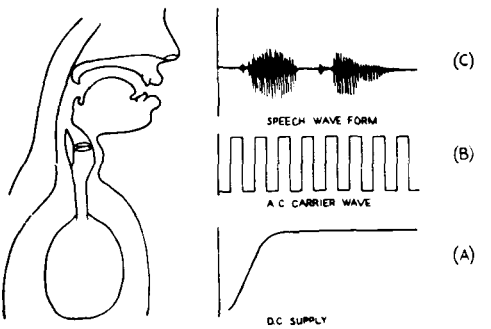


FIG. 2.

The physiological speech mechanism and its physical output. (A) D.C. power supply from the lungs; (B) A.C. carrier wave generated by the larynx; (C) Speech wave-form resulting from modification of the carrier wave by the articulatory mechanism.

And in the contrary sense, as is shown by the work of Eisler (1957), the pause for breath is used as an opportunity for organizing the subsequent stretch of language material, so that there is here an inter-action between the organization of language and of speech. A second factor which influences the point of inspiration is the degree and the kind of emotional tension in the speaker. This factor, too, has been explored in the experiments made by Eisler (1955-56) who has shown that a high degree of emotional tension, if it is freely expressed, leads to an increase in the frequency with which a breath is taken during speech whilst emotion that is felt but repressed leads to a reduction in the frequency of inspiration.

The air provided by the expiratory phase of breathing forms the source of power for the acoustic events in speech and psychological factors, both linguistic and of a more general kind, determine the moments at which this supply of power is to be renewed. The outgoing air forms, as it were, a direct current supply of power, indicated in Fig. 2 by graph (A). In order that the speech movements may give rise to acoustic events, this D.C. supply must at some stage be converted into alternating current and the mechanism which performs this operation is the larynx. The repeated

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

closing of the vocal cords during speech interrupts the steady flow of air and converts it into a series of repeated puffs or pulses. It is not possible even now to say from direct evidence what is the exact shape of these pulses, but in an idealized system of this kind they would have the form shown in graph (B) of Fig. 2. At each moment when the cords open, there is a sudden rise of pressure in the air space just above the cords; pressure remains high during the open phase of the cords and drops suddenly when the cords are closed. The frequent repetition of this cycle of operations produces a series of pulses of approximately rectangular wave-form. This pulse wave train forms the carrier-wave of speech and is the acoustic basis for the sounds of speech. The frequency of vocal cord vibration is under the control of the speaker over a wide range of frequency. In a male speaker with a low voice, the lower limit of frequency may be about 80 c.p.s.; the upper limit in a high-pitched female voice may be 350 to 400 c.p.s. Variation in this frequency largely determines the intonation pattern of the speech. In English and other languages, intonation is associated with the emotional overtones which the speaker adds, usually quite unconsciously, to the factual content of his utterances. A high degree of emotion will generally call up the highest part of the speaker's frequency range, though there are some emotions which cause him to reduce his range temporarily to the very lowest portion.

The amplitude of the carrier-wave supplied by the larynx is not constant. The difference between speaking very loudly and very softly is mainly a matter of change of amplitude in the carrier-wave so that there are long-term changes of amplitude as the speaker proceeds from one group of utterances to another. In addition to this, there are much more frequent changes of amplitude which are due to the syllabic structure of speech. The syllable is a unit which is constantly made use of in discussions of speech and language—it is felt intuitively to be a necessary unit, but it is one of which it is very difficult indeed to formulate a logical definition. It is fundamentally a unit of speech movement, as is made clear by the work of Stetson. The amplitude of speech wave motions reaches a peak once in every syllable and this rise and fall of amplitude is brought about by two factors. The first is a modulation of the carrier-wave amplitude produced by a secondary action of the respiratory muscles. These muscles afford a D.C. supply of power during each utterance, but in addition they vary the amplitude of the power supplied and provide as it were a small surge of power with each syllable. This in itself modulates the amplitude of the carrier-wave coming from the larynx. The second factor in syllabification is the action of the articulatory mechanism—that is, of the tongue, soft palate and lips—which also modulates the amplitude of the carrier-wave. The air passage above the larynx may be more or less widely open during the stream of speech sounds. In every syllable there is a period during which the air passage is wide open

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

and this is preceded and followed by a stretch during which the passage is partially or completely closed. Thus every syllable consists of a middle segment of high amplitude and marginal segments of low amplitude. The action of the articulatory mechanism here reinforces the syllabic action of the respiratory muscles. At some points in the sequence, the carrier-wave from the larynx is modulated 100 per cent., that is to say its amplitude is reduced to zero, as it is in the middle section of graph (C), Fig. 2. The syllabification thus carried out at the physiological level is correlated with the succession of consonants and vowels which form the phonemic sequence at the linguistic level. High amplitude segments correspond to vowels and low amplitude ones to consonants. Stretches during which the carrier-wave is switched off altogether correspond to voiceless consonants.

The syllabic rate varies obviously from speaker to speaker and with the speed of talking. It has also been found that mean syllabic rates differ from language to language. Even so, the overall range of variation is not tremendously wide, the rates varying from about 4 to 7 a second. The incidence of syllables in time constitutes the rhythmic pattern of speech and this again varies from language to language. English has in one sense a complicated rhythmic pattern, since English syllables are uttered with different degrees of force or stress and a speech sequence consists of some strongly stressed syllables accompanied by others more weakly stressed. For a given rate of speaking, the strong syllables tend to fall at regular intervals in English, but the number of weak syllables following each strong beat varies considerably in different sequences. To take an extreme example, in a sentence such as this: *'Evidently the very 'last 'person he ex'pected to 'find there,* there is a wide range of variation in the number of syllables in each foot. The single syllable *last* forms a complete foot; other feet contain from two up to seven syllables, for the first foot in the sentence contains one stressed syllable and six weaker ones. Yet all the feet occupy approximately the same amount of time. It is clear that the mechanism responsible for syllabification in speech must be performing most complex timing operations in order to control the duration and the force of the syllables in accordance with the requirements of the rhythmic pattern.

The other essential function of the articulators is to modify the frequency content of the speech sound-waves. A pulse wave such as the one postulated as the output of the larynx consists of a fundamental with a very extended range of harmonics. The air spaces above the larynx act as a series of resonators which effect further modifications of the carrier-wave by reinforcing certain of its harmonics and attenuating others. Three main cavities are involved in this action—the pharyngeal space behind the root and back of the tongue, the front mouth cavity, under the hard palate, and the nasal cavity. The movement of the articulators, of

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

which the tongue is the most active, maintains a more or less continuous modification of the shape of the cavities during speech. At times the pharyngeal cavity is large and the front mouth cavity small, at others the reverse is true; sometimes the nasal cavity is coupled by an air path with the pharynx, at others it is separated by the raising of the velum and the closing of the air passage at Passavant's cushion. Although the articulators are in continuous motion during speech, there are nonetheless short periods during which something approaching a steady state from the acoustic point of view is reached. These periods correspond with the vowel articulations. During the high amplitude phase of a syllable, the resonating cavities retain a particular disposition for a short time and reinforce particular

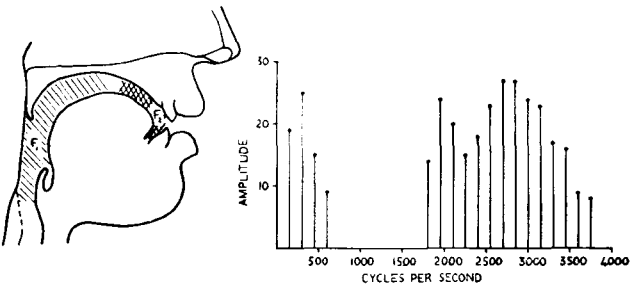


FIG. 3.

Disposition of the resonating cavities for the vowel /i/ and the resulting spectrum.

bands of frequencies. If the carrier-wave consists of a long series of harmonics, all of considerable amplitude, and the carrier is applied to a system of coupled resonators such as the mouth provides during speech, then the resulting wave will have a spectrum in which certain frequency regions are reinforced and most other components are considerably attenuated. High amplitude components produced in this way are called formants. The spectrum shown in Fig. 3 might be the result of resonating cavities such as are shown on the left of the figure for the vowel /i/. Here the tongue is high and forward in the mouth, the space behind the root of the tongue is comparatively large and the front mouth cavity is small. The pharyngeal space gives rise to the lower formant, generally referred to as F_1 and the mouth space to the second formant, F_2, and these formants appear as peaks in the spectrum at 300 c.p.s. for F_1 and 2,700 c.p.s. for F_2. The movements that the tongue makes during speech constantly re-tune the resonators to different parts of the frequency range and thus modulate the carrier in different ways to produce the formants for the sequence of speech sounds.

The consonants which occur between the syllabic sounds are realized in the speech sequence by movements which give rise to noises, noises of the explosive kind, as in /p, t, k, b, d, g/, or noises of the sibilant class like

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

/f, v, s, z/, etc. These consonant movements have two effects on the formant sequence; they establish independent formants for the consonants themselves and they affect the way in which the formants build up to and decline from their steady state positions for the syllabic sounds. Both these effects are important from the point of view of the listener's recognition of consonants, as will be seen later.

From the purely acoustic viewpoint, the articulatory movements produce a continuously changing wave-form in which the fundamental at each instant is determined by the vocal cord frequency, and the wave-shape

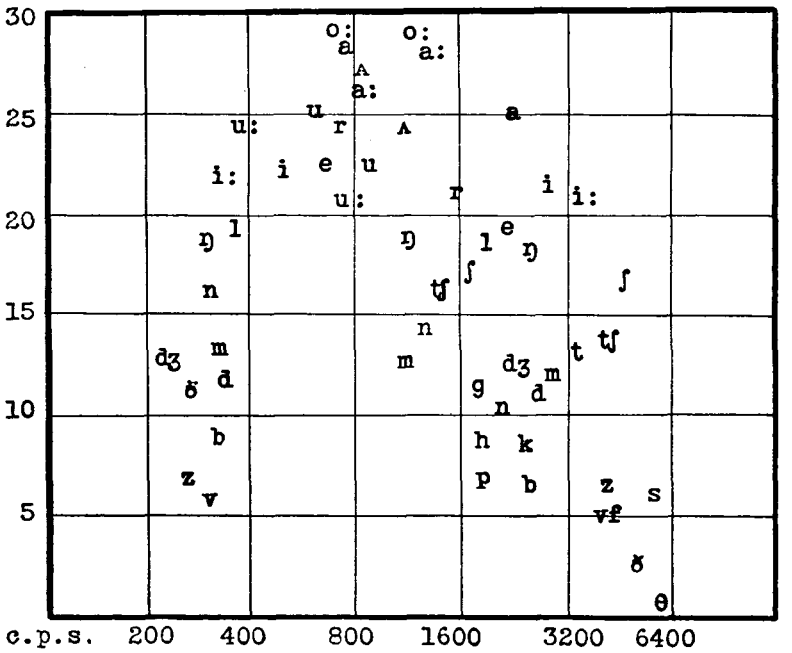


FIG. 4.
The frequency and intensity of the formants for English speech sounds.

depends upon the momentary amplitude modulation of the carrier-wave. A typical stretch of speech wave-form was shown in graph (C) of Fig. 2. The syllables here form the word *marking*. The two vowels correspond to the high amplitude stretches and each has a different wave-form. The consonants also show differences in wave-form, the voiced consonants having a periodicity which comes from the carrier-wave. By analysing a sufficient number of speech wave-forms, it is possible to determine typical frequency positions for formants corresponding to all the sounds in a language. Fig. 4 summarizes this information for English sounds, the data being taken from Harvey Fletcher's measurements. The horizontal axis gives the frequency, the vertical axis the relative intensity of the formants.

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

The frequency of the formants for a given sound is shown by the position at which the symbol representing the sound is printed. The position for F_1 and F_2 is given for all the vowels and consonants and for the nasal sounds the third formant is also shown. Syllabic sounds naturally have high amplitude formants and hence the vowels all appear at the top of the figure, the formants ranging in frequency from 300 c.p.s. up to 3,000 c.p.s. The consonant formants are of less amplitude and are distributed throughout a wide frequency range. A number of voiceless consonants have their upper formant in the highest part of the range, between about 4,000 and 6,000 c.p.s. As the stream of speech proceeds, therefore, the speech movements give rise to an ever-changing pattern of formant frequencies formed by a shifting of the components up and down through the range shown in this figure.

The Feed-back Control of Speech Movements

Before leaving the discussion of the transmission phase of speech communication, it is necessary to examine the ways in which the speech

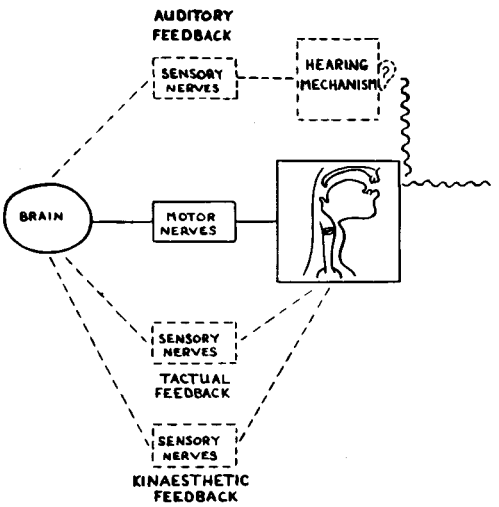


FIG. 5.
The feed-back circuits used in the control of speech movements.

movements are controlled in the speaker. Speech like other human activities is controlled by means of feed-back circuits; that is to say, that during the production of any speech sequence, information is constantly relayed to the brain concerning the progress and the effect of these movements and this information is used in the control of subsequent actions. The feed-back circuits which are important in speech are shown in Fig. 5. The two main circuits are the auditory and the kinæsthetic feed-back circuits, but there is some contribution, too, from tactual feed-back. By far the most

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

important of these is the auditory feed-back loop. The wave-motions set up by the speech movements are transmitted not only to the listener, but also to the speaker's own hearing mechanism and thence to the speaker's brain. Here the information thus conveyed plays an important part in the control of the very finely adjusted movements that speech requires. During a speech sequence, the tongue, for example, is in continuous movement; at one moment it may be drawn up and back in the mouth cavity, then move down to the middle of the mouth, the tongue tip and the sides of the tongue may then rise to make contact along the whole of the upper teeth ridge, and so on. The distance that the tongue moves, the speed at which it moves, the exact moment at which it changes direction—all these are determined largely on the basis of information supplied by the auditory feed-back. The fact that the complex and exact timing in speech is very much dependent on auditory feed-back is demonstrated by the effects of "delayed feed-back" (Lee, 1950; Black, 1951). If a speaker is made to speak into a microphone, his speech can be fed to his ears through a pair of telephone receivers, and this speech can be made so loud that he no longer hears anything of the bone-conducted sound on which the auditory feed-back depends to a considerable extent. Having obtained this condition, a time delay can be introduced into the artificial feed-back so that the speaker hears his own speech sounds, not at the moment when he expects them, but a fraction of a second later. If the period of delay is about 1/10 second the effect on most speakers is very dramatic; an artificial stammer is produced and the speaker in some cases finds it impossible to proceed with speech movements at all. This shows that the speech mechanism is very sensitive to interference with the time sequence of events.

The second main feed-back circuit in speech is the kinæsthetic feed-back by which information is sent to the brain concerning muscle states and hence of muscle actions. This feed-back is in continuous use in everyday life since all our movements depend upon it. It is only necessary to point out that speech movements are among the most finely co-ordinated movements that man makes, and their control depends very much on the continuous flow of information from the muscles themselves to the brain. In addition, the brain receives some information from the tactual feed-back, since parts of the tongue, the roof of the mouth and the lips are all capable of tactual sensation. As far as one can judge, however, this particular circuit does not play a very great part in speech.

In principle, then, control of speech movements is through a system of feed-back circuits. The nerve connections which constitute these circuits are very complex indeed and there is a great deal about them that is not yet known. The control of breathing and phonation, on the one hand, and of articulation, on the other, are to some extent independent of each other. Breathing and phonation are neurally controlled at a lower level of integration than articulation, and are more directly subject to the influence of

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

emotional states. The voice quality the speaker uses and the intonation pattern of the speech are both aspects of phonation and are the exponents of emotion in speech. Emotional disturbances which affect speech generally interfere with phonation rather than articulation. With regard to the feed-back control of phonation, it is clear that the auditory path is the most important in determining the mean amplitude and frequency of the carrier-wave. Every speaker tries to maintain the loudness of his speech at a level which gives a satisfactory signal/noise ratio. If the noise surrounding him is increased, he will raise the power of his voice in an attempt to recover the right ratio and will continue to do so even beyond the point at which it becomes physically impossible for him to compete with the noise level. In the control of frequency, too, the auditory feed-back is very important. Experiments by Deutsch (1957) have shown that obliteration of the auditory feed-back by very loud masking noise makes it impossible for a subject to sing a sequence of notes in tune, or to maintain a note at a required frequency. This effect does not show up in speech, as the tolerances for vocal cord frequency are very wide in this case. Speech intonation patterns depend on change in vocal cord frequency and not on the sustaining of a constant frequency; the direction of change is all-important and the amount of change is relatively unimportant. In these conditions, vocal cord frequency can be satisfactorily controlled through the kinæsthetic feed-back, and this is in fact what a speaker does who is obliged to speak in a very loud noise.

Recent neurological work on phonation, by Husson (1950) and his co-workers, has somewhat changed the general picture of the control of phonation. It was for a long time believed that the function of the laryngeal nerves in speech was to control the physical characteristics of the laryngeal tissues and that the vibration of the vocal cords was a result of air pressure applied from the lungs. In France recently experiments have been carried out on patients undergoing laryngectomy and these indicate rather that vocal cord vibration may be controlled by nerve impulses cycle by cycle. In cases where the frequency of vibration is too high to be followed in one nerve fibre, it is suggested that there is alternation in different fibres.

The neural control of articulation is more complex than that of phonation and calls for the co-ordination of tongue, palate and lip movements. What we might call the "programming" of articulatory movements seems to be effected in syllabic units; that is to say, the brain organizes its instructions to the motor system so as to produce complete syllable movements and not just successive speech sounds. For example, the movements for a syllable such as *strike* are organized as a whole; once embarked on this syllable the articulators cannot be redirected to utter the syllable *speak*, or even *stripe*. In ordinary speech, when a speaker wishes to correct a mistake in articulation he is compelled to begin the syllable again;

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

very often, of course, he will begin a whole word or a group again, but he cannot certainly make a correction in mid-syllable. The chest muscles receive an operating instruction at every syllable and the chest and the articulatory muscles are to some extent locked together. Furthermore, time delay in the auditory feed-back has its maximum effect when the delay is something less than the average syllable duration; in other words the most serious disruption of speech movements occurs when the timing is interfered with in the course of the syllable.

The control of speech movements depends on the continuous "monitoring", as it were, of the muscle actions through the auditory and the kinæsthetic feed-back loops. The co-ordination of information supplied by different feed-back circuits is one of the vital brain activities during speech. It is important to notice that, once we have learnt speech, we can maintain the flow of speech movements without the aid of the auditory feed-back. A speaker placed in a noise so loud that it completely masks the sound of his own speech, can continue to speak and his movements are not abnormal except in so far as he inevitably tries to produce enough power to overcome the masking. The sequence and timing of movements remain normal and they are in this case based on the kinæsthetic and tactual feed-backs. The situation is very different at the stage of learning to speak; the auditory feed-back is of the utmost importance during the time when the habits of speech movement are being formed. A child who is born deaf will not learn speech—and hence will not learn his native language—unless we take special measures to compensate for deficiencies in his auditory feed-back. It is most important in such a connection to use and develop to the fullest extent any hearing that the child has, and in addition great use has to be made of touch and vision in order to develop the control of speech movements, mainly through the kinæsthetic feed-back path. Learning to speak, from one point of view, therefore, consists in building up a large memory store of kinæsthetic patterns on which the speaker calls in future speech activity, and which, if need be, will enable him to control speech movements even in the temporary absence of auditory feed-back.

The Reception of Speech

The feed-back circuits in speech form a very complex system, operating at different levels and so interconnected as to provide co-ordination of breathing, phonation and articulation. The speech movements have acoustic effects which are transmitted to the listener's ear and the rest of the speech chain is concerned with recovering from the acoustic signals the language units which constitute the original message.

The listener must, if he is to understand the message, possess a memory store of linguistic information similar to that held by the speaker—this indeed is what we mean when we say that two people have a common

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

language. There is in the brain of the listener a knowledge of linguistic units and of the ways in which they combine similar to that pre-supposed in the brain of the speaker. The most important factor in the reception of speech is the listener's knowledge that whatever message he receives, it must be made to fit into the patterns that belong to his language. As Wiener has put it, "the most important information we can have about a message is that it makes sense." Once the listener has established that the message is in a language with which he is familiar, he will then make every effort to "make sense" of the message, no matter how distorted or unintelligible the sounds arriving at his ear. The "making sense" consists in transforming the speech wave-motions back into linguistic units.

The reception of speech depends on a combination of two processes, one based on perception of the incoming sound-waves and the second on the listener's *a priori* knowledge of the language. The action based on perception may be called primary recognition; this means the action by which a listener distinguishes sounds *per se*. For example, an English listener can distinguish the syllable *pa* from the syllable *ta*, even with no verbal context to help him. In response to the question "What do you hear?" he is capable of replying "The first syllable begins with *p*, and the second with *t*."

In primary recognition, the listener is already making use of linguistic classes. He receives the first syllable and assigns the first sound he hears to the phoneme /p/; in the second syllable he assigns the first sound to the phoneme /t/. The distinction is based on perceptual differences between the two sounds, but perception of speech sounds is itself subject to the organization imposed upon it by the language system. In English, the listener has, as it were, 40 phoneme boxes and every incoming sound is put into one of these boxes. The process by which this differentiation is carried out still requires a good deal of investigation but much interesting material has already been provided by experiments carried out mainly by Cooper and his colleagues (1952).

The influence of linguistic organization on perception has been demonstrated in a series of experiments by Harris and others (1956), using synthesized speech sounds. It is possible to produce synthetically a series of artificial syllables in which the first sound is changed by very small steps from /b/ to /d/ and from /d/ to /g/, i.e., through the series /ba/, /da/, /ga/, for example, but in such a way that /b/ gradually becomes /d/ and /d/ changes gradually to /g/. These synthesized syllables were used in two types of listening experiment. First, listeners were asked to identify each of the syllables in the series as /ba/, /da/ or /ga/. As we should expect, there were three stretches in the scale of continuous change over which the listeners agreed completely in their judgments; for one stretch everyone agreed the syllable was /ba/, for second, that it was /da/, and for a third, that it was /ga/. Between these stretches, the degree of agreement between

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

the listeners first decreased until the judgments were about equally divided and then increased again as the next plateau was approached. This experiment, therefore, showed up discontinuities in a perceptual scale that are equivalent to phoneme boundaries. The next experiment consisted in establishing variations in differential threshold for this same series of syllables; that is to say, a measure was obtained of the listeners' sensitivity to change along the scale. It was shown quite unequivocally in these experiments that the listeners could detect much smaller changes near the phoneme boundary than they could over the middle stretches where, in the previous experiment, there was shown to be no doubt what the phoneme was.

It is clear then that even in primary recognition, our perceptions are very much subject to the organization imposed on them by the long process of learning our native language—or indeed any other language that we speak. The perceptual side of recognition must obviously depend upon certain acoustic features of the sound waves, but as we have just seen, the listener's sensitivity to these features is affected by learning. Two other cardinal facts must be noted about the relation of acoustic features to phonemic recognition. First, these acoustic features reside in relations between physical quantities and not in absolute values, in relations of frequency, of intensity, of harmonic structure, of time patterns of acoustic change. To take a simple example, the vowel formants in a child's speech are not the same as in a man's, but we can recognize both sets of sounds because of the relations between the vowels within one set. In a similar way, the noise that constitutes a /k/ varies in different contexts, but we have no difficulty in recognizing the sound both because of its relation to a following vowel and to the sounds /p/ and /t/ with which it is most frequently commutable.

The second fact is that there are always a number of cues that a listener can use in recognizing any one sound—he is never dependent on a single invariable physical cue. In this field, too, most interesting work has been done by Cooper and his colleagues. They have shown that plosive consonants can be recognized from the frequency of the noise produced by their explosion, but they can be recognized too from the way in which the accompanying vowel formants change with time. Similarly, fricative consonants, /f/, /v/, /th/, /s/, /z/, /sh/, etc., can be recognized from the frequency of the sibilance and also from the change in the vowel formant; other experiments by Denes (1955) have shown that the distinction between /s/ and /z/ for example can be established by the relative duration of a preceding vowel.

Primary recognition is in itself, then, a very complex operation which we have only recently begun to understand at all. The role played by primary recognition has, however, generally been rather over-estimated. There is another factor in the reception of speech which is equally strong

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

and which can in fact override decisions made by primary recognition. This factor is the statistical knowledge of the language which every listener carries in his brain. He knows what sequences of phonemes, morphemes and words are possible in the language. Given any string of phonemes, he knows, without being aware of it, what is the probability that any particular phoneme will be the next one and he uses this information all the time in his reception of speech. The fact is that the listener probably *hears* no more than about a third or a half of the speech sounds which reach

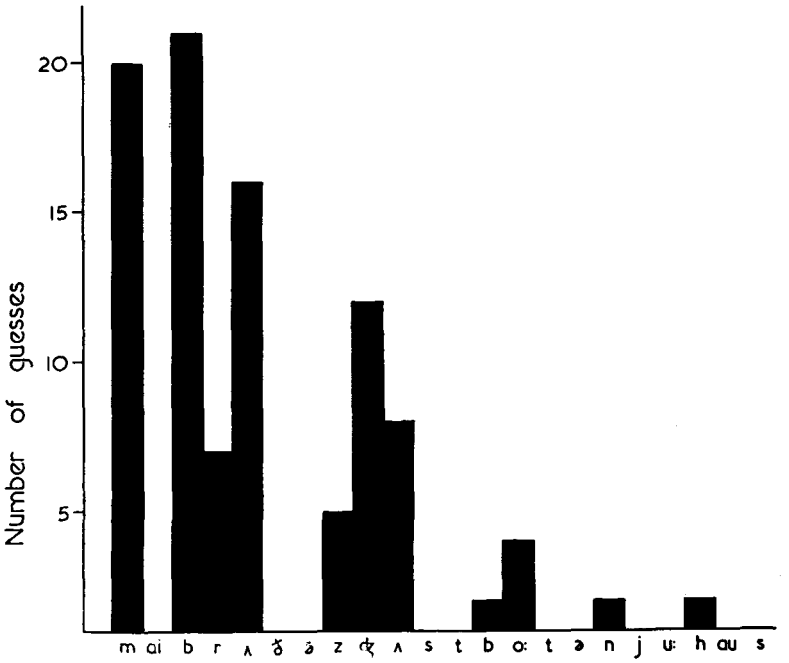


FIG. 6.

The result of an experiment that demonstrates the linguistic knowledge accessible to the listener. The figure shows the number of guesses required by a subject in guessing successively the phonemes in the sentence "My brother's just bought a new house".

him, and he is able to supply the rest of each message by using his knowledge of linguistic restraints. A demonstration of the existence of this knowledge can be obtained by a modification of an experiment made by Shannon (1948), one of the originators of information theory. An English sentence is selected, and a listener is asked to guess what is the first sound in this sentence, without knowing anything about the sentence or its context. He needs a number of guesses, though he is likely to guess the right sound without using up the whole list of forty phonemes. He is told as soon as he guesses the first sound correctly, and then goes on to guess the second sound, and so on through the sentence. The results of this experiment are shown in Fig. 6. The selected sentence was: *My brother's just bought a new*

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

house; a transcription of the sentence is shown along the horizontal axis. The subject took 20 guesses to discover that the first sound was /m/, but, as it happened, he guessed /ai/ with his first guess—in other words, knowing the first sound was /m/, he immediately guessed that the first word was *my*. He then took 21 guesses to get /b/ in *brothers*, and so on. The important feature of this result is that the number of guesses required shows a general decrease as he proceeds through the sentence; in other words, he becomes more and more certain what the sound must be. Altogether, out of 22 sounds, the subject guessed 11 correctly with his first guess.

This knowledge of what is coming next, or is likely to come next, is all-important in the reception of speech and when we listen we are all the time filling in gaps by using our linguistic knowledge. Primary recognition provides, as it were, a scaffolding on which we erect linguistic patterns that conform to the demands of the language structure. The process begins at the phonemic level and proceeds, probably by a series of scanning operations, to other linguistic levels. The ear is presented with a continuous sound-wave which the brain segments into discrete language units. The first stage in this process is the conversion of the input into patterns bounded by psychological dimensions, of which four are basic—quality, pitch, loudness and length. An incoming pattern is compared with a stored repertory of patterns which is scanned in the process of primary recognition, and as a result, a phoneme is recognized, let us suppose the phoneme /i/. The next pattern perhaps is recognized as /t/, and this process is continued at the phonemic level. At the same time, on the next level, the morphemes are being scanned for correspondence with phoneme sequences that result from the first scanning, so that in our example, *it* would be recognized as a morpheme. This kind of process is repeated at the word level and finally at the sentence level.

One feature of this serial scanning process must be stressed. Just as in the case of the control of speech movements, there is a continuous feeding-back of information from one level to those below it, so that the process consists in feeding forward the results of scanning and also feeding back information as to the constraints which operate. This mechanism affords the means of error correction and also greatly shortens the scanning operations. Suppose that the phoneme level has fed forward two units: /l/, /e/. These are being scanned for coincidence with a morpheme and at the same time the next pattern is being dealt with at the phonemic level. There are a considerable number of phonemes which cannot occur after the sequence /l/, /e/. Since /e/ can occur only in a closed syllable, for example, no vowel can follow it; a number of consonants, too, are impossible—/h, j, w/ and /r/ cannot close a syllable, so that only about 20 out of the 40 odd phonemes are left to be scanned because of the linguistic information fed back to the scanning mechanism. Suppose that the next pattern is recognized as /t/, and this information is fed forward to the morphemic level

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

where the sequence is recognized as a morpheme, *let*. This immediately reduces materially the number of morphemes which must be scanned to fill the next morpheme position, and consequently the phoneme recognitions are further restricted since they must not be at variance with the possible morpheme sequences. This process then continues at all levels. *let* may be followed by *me* and this gives two words on the word level. If primary recognition led the listener to hear *sink* as the next part of the sequence, there would be a suspension of decision at the morpheme level. *sink* is a possible morpheme after *let me*, but it is an improbable one; if the further continuation were the word *about*, feed-back from the word level would impose a correction of the phoneme /s/ in *sink* to the /th/ phoneme and it would be clear that the morpheme in question was *think*, a decision that would be clinched by the occurrence of *it* in the next position, completing the sentence *Let me think about it*.

This review of the speech reception process has not so far taken into account one important factor: the part played in the listener by the recall of kinæsthetic images. Reception and recognition of speech sounds depend on the recall not only of auditory patterns, but of patterns of motor activity which form the listener's experience as a speaker. Very little is yet known about this aspect of speech reception, but its importance was recognized and stressed by Hughlings Jackson. Primary recognition at the phonemic level no doubt involves some measure of appeal to the listener's own motor habits, and it is probable that the recognition of other features in a speech sequence, such as stress and rhythmic patterns, for example, depends even more upon this factor.

Nearly all the speech and language processes that have been described in this paper take place at a brain level where the listener is unaware of them. On the transmission side, the linguistic organization of the message and the operating instructions to the speech mechanism, on the reception side, the scanning operations, the feeding-back of statistical linguistic information, the connections with motor images, all form part of the brain activity in speech. It will doubtless be a long time before we begin to understand how the actions are carried out, but the basic structure of the system has already been suggested by Hughlings Jackson, whose observations and reflections will bear interpretation in the terms that have been used in this paper. It would appear that the linguistic organization is developed on both the right and the left side of the brain. The links between the language level and the motor speech mechanism (including the feed-back circuits) are developed, in the right-handed person, only on the left side. The right side normally specializes in the recognition and reception of speech, operations which are concerned mainly with language. The right side, however, includes the necessary anatomical structures for the motor activity of speech and these are presumably called upon in the re-education of an aphasic patient whose speech has been destroyed by a

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press

brain lesion. It may well be indeed that the contribution made by the recall of motor patterns comes also from the right side. In this case the real difference between the right and left side, as has been suggested above, would lie in the absence, on the right side, of developed paths for linking linguistic organization with actual speech movements. This view would at the same time be consistent with the frequently repeated observation that patients suffering from loss of speech are often able to produce some purely automatic speech sequences. Naturally the brain activity in speech is much more complex than this over-simplified view of the functions of the hemispheres suggests. The anatomical connections between the two sides are very complicated indeed, but so, too, are the inter-relations between speech and language. It is possible that by a thorough-going attempt to correlate our knowledge of speech and language with the study of disturbances of speech of all kinds we might begin to fill in some of the gaps in our understanding of what is after all one of the most complex of human activities.

REFERENCES

BLACK, J. W. (1951) *J. Speech Hear. Dis.*, **16,** 56.

DELATTRE, P., COOPER, F. S., LIBERMAN, A. M., and GERSTMAN, L. (1954) *J. acoust. Soc. Amer.*, **26,** 137.

DENES, P. (1955) *J. acoust. Soc. Amer.*, **27,** 761.

DEUTSCH, J. A. (1957) Unpublished communication.

FLETCHER, H. (1953) Speech and Hearing in Communication, 2nd edit. New York.

GOLDMAN-EISLER, F. (1957) Paper read to the Experimental Psychology Group, London.

——— (1955) *Brit. J. Psychol.*, **46,** 53.

——— (1956) *Brit. J. med. Psych.*, **29,** 35.

HARRIS, K. S., LIBERMAN, A. M., HOFFMAN, H., and GRIFFITH, B. C. (1956) *J. acoust. Soc. Amer.*, **28,** 760.

HUSSON, R. (1950) Thèse Fac. Sci., Paris.

——— (1952) *Ann. Oto-laryng. (Paris)*, **69,** 124.

JACKSON, J. H. (1932) Selected Writings of John Hughlings Jackson, Vol. II, London.

KELLER, H. (1903) The Story of My Life, London.

LEE, B. S. (1950) *J. acoust. Soc. Amer.*, **22,** 824.

LIBERMAN, A. M., DELATTRE, P., and COOPER, F. S. (1952) *Amer. J. Psychol.*, **65,** 497.

DE SAUSSURE, F. (1949) Cours de linguistique générale, Paris.

SHANNON, C. E. (1948) *Bell System Tech. J.*, **27,** 379 and 623.

STETSON, R. H. (1951) Motor Phonetics. 2nd Edit., Amsterdam.

https://doi.org/10.1017/S0022215100052002 Published online by Cambridge University Press