H. MOL

Fundamentals of Phonetics

I

MOUTON & CO. - THE HAGUE

FUNDAMENTALS OF PHONETICS

JANUA LINGUARUM

STUDIA MEMORIAE
NICOLAI VAN WIJK DEDICATA

edenda curat

CORNELIS H. VAN SCHOONEVELD

STANFORD UNIVERSITY

SERIES MINOR

NR. XXVI

1963

MOUTON & CO · THE HAGUE

FUNDAMENTALS OF PHONETICS

I: The Organ of Hearing

by

H. MOL

UNIVERSITIES OF AMSTERDAM AND LEIDEN

1963

MOUTON & CO · THE HAGUE

The chapters III and IV of this book are based on the author's contribution to a paper by H. Mol and E. M. Uhlenbeck: "Hearing and the Concept of the phoneme", *Lingua*, VIII, 2 (1959), p. 161–185.

Printed in the Netherlands by Mouton & Co., Printers, The Hague.

TABLE OF CONTENTS

INTRODUCTION

When listening to speech the listener is expected to interpret the acoustic cues the speaker produces. The notion "acoustic cue" must be rigorously defined as a physical phenomenon able to elicit a certain characteristic pattern of nerve impulses in the nervous system. This pattern of nerve impulses need not generally be regarded as a foolproof description of the speech sounds in question. On the contrary, experience shows that quite often, these nerve impulses merely *help* the listener in recognizing the words intended by the speaker without actually describing them. Recognition is the result of a decision taken by the brain. This recognition may be correct or not. The brain does not base its decisions merely on the acoustic cues. As a matter of fact it cannot do so because, especially for vowels, these cues are ambiguous. The brain must marshall its knowledge of the language, the situation and the speaker's vocal habits in order to be able to fit the acoustic cues into the frame work of limited possibilities. This is in fact a playful activity, a game. We shall, in this monograph, try to indicate which type of acoustic cues the ear is able to transform into nervous activity, basing ourselves on recent data on the physiology of the ear and the nervous system and the outcome of experiments with models of nerve elements.

In order to enable the reader fully to appreciate the limited possibilities of the organ of hearing we devote our first chapter to a short glance at the fundamentals of the theory of sound.

We shall avoid the use of formulas as much as possible. In a special appendix the limitations of Fourier analysis are discussed in some detail. The study of this appendix, however, is not necessary to understand the main chapters of this book.

We refrained from describing speech sounds in terms of sinusoidal

vibrations. The reason for this is our conviction that Fourier analysis, also often called frequency-analysis, is a mathematical tool extremely ill-fitted to the purpose of explaining the mode of action of a sense organ like the ear that is simply alive with non-linear mechanisms. In audiometry, however, pure tones (i.e., sinusoidal vibrations) can be used to advantage for diagnostic purposes. The location of pathological changes in the ear with the aid of pure tones, however, is one thing, whereas discovering the way in which the ear is helpful in classifying speech waves, is quite another thing.

We think it is an unwise policy to fill the mind of the reader from the start with the doctrine of decomposing sound curves into sinusoidal partials as if this procedure were the only and even necessary tool for research. It is our experience that most people, after having adopted this narrow-minded way of thinking, are not able to free themselves from it, and are consequently led into a blind alley when faced with the problem of discovering the real nature of the auditory mechanism.

FUNDAMENTALS OF THE THEORY OF SOUND

Hearing is based on the interpretation of small variations of the barometric pressure. Most people are already more or less acquainted with the gross variations of the atmospheric pressure they can read on the barometer. The classic barometer consists of a glass tube, closed at one end and open at the other, which is

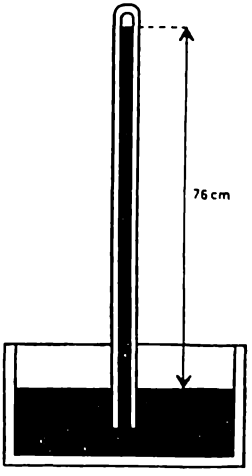


Fig. 1. The classic barometer.

filled with mercury and immersed in a vessel likewise filled with mercury. The pressure of the air on the surface of the fluid in the vessel pushes the mercury into the tube until, under average conditions, the height of the column equals 76 cm.

In meteorology, and also in every day life, one is interested in the gross and comparitively slow changes of the barometric pressure which accompany or even precede, changes of the weather. Variations of several centimetres are possible.

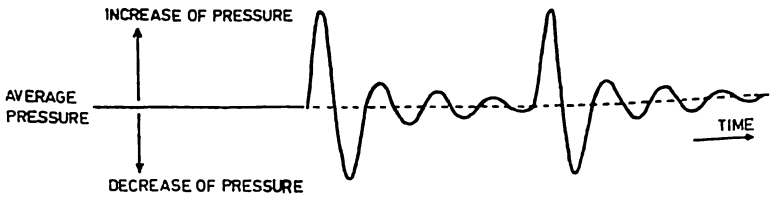


Fig. 2. Example of a sound curve.

In acoustics, however, very small and extremely rapid variations of the atmospheric pressure play a part. Even in loud speech the barometric pressure undergoes changes of about $\frac{1}{1\,000\,000}$ or 10^{-6} of its average value. It stands to reason that the classic barometer (and also its more modern versions) cannot indicate these extremely feeble oscillations, the more so as the inertia of the heavy fluid column does not permit oscillations at a high frequency. The fact that man labels these variations as loud betrays the amazing sensitivity of the human ear.

Instruments capable of registering the small pressure variations occurring in sound waves are called microphones. In fact the human ear is mainly a microphone or, even better, a collection of microphones. In order to understand the mode of action of a microphone we must introduce the notion "sound curve".

A sound curve is a graph portraying the way in which the atmospheric pressure varies from moment to moment. Mathematically speaking it is a time function. "Seeing is believing" especially holds in acoustics: the investigator can often extract more valuable information from a graph than he can do by listening. Sometimes it is just the other way, but in my view this is caused by the fact that he then looks at the wrong curves.

In fig. 2 an example of a sound curve is given. It shows how, in this particular case, the pressure of the air oscillates around its mean value. Only oscillations are interesting to the human ear. On the horizontal axis time is plotted, on the vertical axis air-pressure is indicated. The zero-line represents the average barometric pressure around which oscillations take place. In this case, not all the pressure peaks are equally high, the curve has the

typical shape of periodic repetitions of the so-called damped oscillations so common in phonetics.

Though it is customary to show only pressure curves, probably because they can so readily be measured by microphones, they tell us only part of the acoustic events taking place in the sound field.

For a complete calculation of the acoustic vibrations in air we must take into consideration three quantities, that is to say the already mentioned pressure, the velocity of the vibrating air particles, and the density of the air. The air can be considered as consisting of small compressible particles that move to and fro under influence of pressures exerted on them by adjacent particles. When a particle is displaced by its neighbour it is also compressed. We have three unknown quantities: pressure, velocity, and density, which vary with time. In order to determine each of these three we need three equations, in fact three Laws of Nature.

The first law says that the resultant force exerted on an air particle is proportional to the acceleration forced upon that particle.

The second law expresses that though the vibrating air might be compressed or rarefied at no point of the sound field matter can be created or destroyed.

The third law states that the displacements of the air particles take place at such short time intervals that no heat exchange with the environment can occur. This is a way of saying that so-called adiabatic processes are going on.

By combining the three laws either the pressure or the velocity or the density can be calculated by eliminating the two remaining quantities. The result presents itself in the form of a mathematical equation in which only the desired quantity and the time appear. It is called the wave equation. Unfortunately it can only be solved in special, often idealized cases.

As already stated we are most interested in sound pressure because we can directly measure this quantity or trace its time course with the aid of microphones. Sometimes, however, we must also take into consideration the particle velocity in order to understand what is going on. For the sake of simplicity we shall in most cases confine ourselves to sound pressure. Unless otherwise

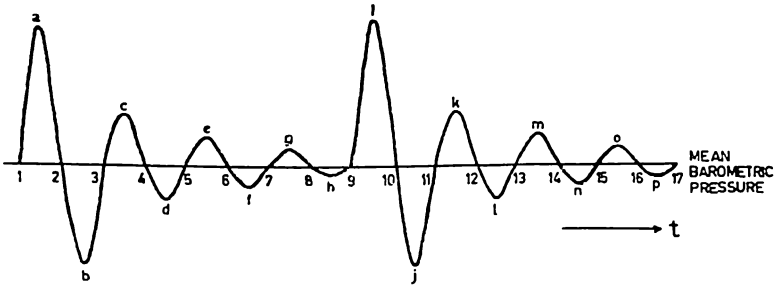


Fig. 3. Example of a sound curve.

indicated, our sound curves refer to the variation of the pressure at the entrance of the ear with time.

It is the very essence of a sound curve that the barometric pressure differs from moment to moment. Fig. 3 shows an example of a vowel-like curve chosen at random.

It represents an oscillation around the mean barometric pressure. The points 1–17 represent the moments when the pressure equals the barometric pressure. These moments are called the zero crossings. As we shall see later on the zero crossings are very important for the human ear by which they can be readily defined.

Furthermore the curve shows peaks, that are relative maxima, indicated by the letters a–p. Not all peaks are equally high: a and i are the highest positive peaks, whereas b and j are the highest negative peaks of this example.

Very likely the ear has no means of determining the individual heights of the successive peaks though by means of a complicated nerve mechanism it will, no doubt, perform some sort of averaging action involving an integration.

It presents to the brain a spatial and temporal pattern of nerve impulses on basis of which a loudness-sensation is experienced.

There is a tendency to ascribe to a sound courve an objective quantity, to be described by only *one* number that is thought to be representative of the vehemence of the acoustic disturbance.

Strictly speaking the choice of such a quantity is arbitrary. When, for instance, the sound pressure is first squared and then averaged

over a certain period after which the square root is taken the result is the RMS (Root Mean Square) value.

When the highest peak in a certain period is considered as representative of the sound curve we call it the peak value of the curve.

In general neither the peak value nor the RMS value of a sound curve can be used to predict the "loudness" of that sound. The value of the psychological notion "loudness" in linguistics is very questionable. Though it is tradition to speak of the dynamic accent the loudness of syllables does not *function* in intonation though differences in loudness are certainly present in speech waves. When, however, by means of a regulating amplifier, one makes all syllables equally strong, this procedure appears not to impede the intonation of the utterance in question.

In phonetics (and also in music) we often encounter sound curves that consist of time patterns which repeat themselves periodically. This is the case in, for instance, vowels. In fig. 4 some examples are given.

The duration of a repeated pattern is indicated by its period T. We might also call it the repetition period T or the periodicity T. It is expressed in seconds, more often than not in milli-seconds, a milli-second being the thousandth part of a second.

Another possibility of describing the periodic character of a curve is stating the number of patterns that occur per second. This number is known as the frequency or repetition frequency or repetition rate of the pattern in question and indicated by the symbol f.

Obviously

$$f = \frac{1}{T} \quad \text{and} \quad T = \frac{1}{f}$$

Formally speaking we may use either f or T as these quantities are inseparately tied together by these equations.

In oscillograms, however, it is the time intervals, the periods, that present themselves to our eyes. Also, when the total duration of a periodic curve is very short, it makes less sense to state the number of patterns per second. Furthermore, in our opinion the

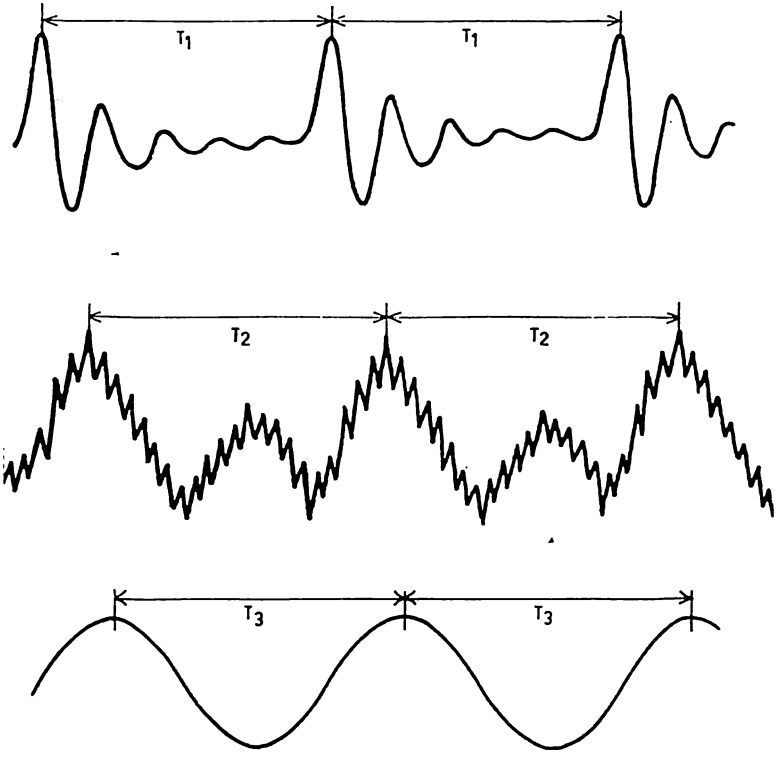


Fig. 4. Examples of (nearly) periodic sound curves.

nervous system is better suited to measuring time intervals than to counting a number of patterns occurring in a certain time interval.

We shall, therefore, in this monograph, describe periodic curves by their periods.

Comparing the examples shown in fig. 4 we see that the curves have different shapes as well as different periods.

Differences in shape betray themselves to the listener as differences in timbre or quality though the exact nature of this process is still obscure.

Differences in periodicity are experienced by the listener as differences in what he calls "pitch". The larger the period T becomes the lower the resulting pitch is called.

Two sounds may be termed equally "high", meaning that their periods are equal, whereas the quality (timbre) of these sounds may be totally different.

Before the Giorgi-system of units became accepted sound pressure was expressed in dynes/cm^2. Nowadays it is expressed in Newton/m^2. It is worth knowing that 1 dyne/cm^2 is equivalent to 0,1 Newton/m^2.

In audiology it has become a tradition to compare two values of the sound pressure not by determining their difference but their ratio. This tendency has developed into the habit of always describing a sound pressure as its ratio to a predetermined reference value.

In acoustics a reference value of

$$P_0 = 2.10^{-5} \quad \text{Newton/m}^2 = 2.10^{-4} \quad \text{dyne/cm}^2$$

has been internationally agreed upon as being near to the auditory threshold. So a sound pressure P is described by the ratio $\frac{P}{P_o}$, a dimensionless quantity.

It has become a tradition not to handle the ratio $\frac{P}{P_o}$ directly but to define the so-called pressure level as $20 \log \frac{P}{P_o}$. The pressure level is expressed in decibels, abbreviated to dB. This procedure is inspired by the fact that the ear can cope with an enormous range of sound pressures the highest of which is $1000000 = 10^6$ times as big as the weakest. In this case the dB-scale yields a comparatively low number, namely $20 \log 10^6 = 120$ dB.

Another excuse for the decibel scale is that when the sound pressure is multiplied again and again by the same factor, say 2 just to take an example, the hearer has the impression that his experienced loudness increases with the same "step" (whatever he may mean by that!). This fact fits into the logarithmic scale because the pressure level increases by the same step, in this case $20 \log 2 = 6$ dB, as is apparent from

$$20 \log 2\frac{P}{P_o} = 20 \log \frac{P}{P_o} + 20 \log 2 = 20 \log \frac{P}{P_o} + 6$$

In phonetics, however, we have seldom if ever use for the dB-scale unless we forcibly introduce it.

The sound curve is the only physical reality with which the ear is confronted. The ear has its own means of reacting to a sound curve, and it produces its reaction in the form of a pattern of nerve impulses that is propagated along the fibres of the acoustic nerve to the higher centres of the nervous system.

The ear as such does not "know" anything about the physiology of the larynx and the vocal tract. It is not conversant with the man-made mathematical methods of solving the differential equations of the vocal tract and the favoured position Fourier-analysis holds among mathematicians, physicists and engineers. The ear reacts as it were automatically to a sound curve because of the physiological limitations of its constituent parts.

There is no harm in intuitively calling the reaction of the ear an analysis or a decomposition as long as we do not burden ourselves with the perplexities and paradoxes of a mathematical method, especially Fourier-analysis, before it is shown that this is absolutely necessary.

THE MICROPHONIC PROPERTIES OF THE EAR

As the nervous system works on an electro-chemical basis it is obvious that we must look for so-called microphonic elements in the ear. A microphonic element is a structure that is able to transform a mechanical motion into an electric phenomenon, if possible an exact electrical replica of that motion. Here it is instructive to make a comparison with electro-acoustics.

The tremendous development of electro-acoustics has, a.o., been made possible by two important inventions, namely, the microphonic element, and the amplifier.

Normally the electrical power the microphonic element can produce is at best equal to the acoustical power derived from the sound field, which is usually very low. An electronic amplifier is able to step up that power in order to drive loudspeakers, oscilloscopes, recorders, analysers, etc.

The world's most wide-spread microphone is the so-called carbon microphone in the telephone subscriber's set. Its microphonic element is formed by a chamber filled with carbon granules. As shown in fig. 5 one of the walls is fixed and serves as an electrode. The opposite wall is movable and also serves as an electrode. It is connected with a diaphragm that is moved to and fro by the sound pressure so that the granular mass is alternately compressed and rarefied. Consequently the resistance of the granules is alternately decreased and increased. As the carbon chamber is transversed by an electricial polarisation current J_o, it produces an electromotoric force e given by

$$e = J_o.\Delta R$$

where ΔR symbolizes the change the movements of the diaphragm bring about in the resistance of the granular mass.

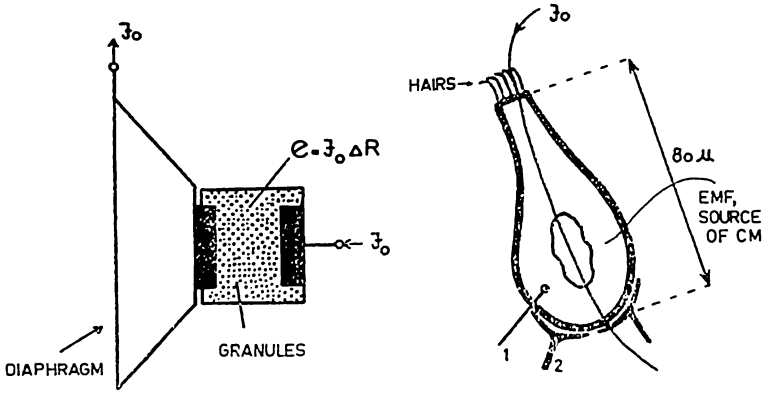


Fig. 5. The resemblance between the carbon microphone and the hair cell. They both amplify.

The electromotoric force e is able to deliver more electrical power into the telephone network than the diaphragm of the carbon microphone derives from the sound field. This extra-power is furnished by the battery that produces the polarisation or feeding current J_o. Consequently the carbon microphone is a microphonic element as well as an amplifier which explains its popularity in spite of its inherent distortion. In commercial telephony, where speech must be transmitted over long lines, high powered microphones are in demand.

The human ear does not house one microphonic element only, but instead some 23,500 in the shape of the so-called hair cells neatly arranged in the organ of Corti, the ear's most important structure which we shall learn to know better in the next chapter.

A hair cell (see fig. 5) is about as big as a red blood corpuscle. It has an oblong shape, its length being about 80 μ (1 $\mu = \frac{1}{1000}$ of a millimeter).

The cell is called a hair cell because it carries some 30 stiff hairs on its head. When these hairs are pulled or bent a microphonic potential (also called cochlear microphonics, abbreviated as CM) is generated in the immediate vicinity of the cell. With the aid of small electrodes one has been able to measure the microphonic

potential in living animals. As far as we know now an electrical polarisation current J_o flows through the hair cell. Manipulation of the hairs causes resistance changes in the hair cell so that an electromotoric force, betraying itself as the microphonic potential, is generated. It has been shown experimentally (10)[1] that CM is proportional to the displacement (that is the excursion from the position of equilibrium) of the organ of Corti. In other words: CM is an exact electrical replica of that displacement. It is also known that CM increases and decreases as J_o is artificially increased or decreased. Also the nervous activity in the auditory nerve goes up and down with J_o.

It is not quite certain, however, whether CM is the direct stimulator of the nerve endings that arborize around the hair cells. It is possible that there is an intermediate chemical process involving acetylcholine.

Apart from the microphonic potential (which is an alternating voltage) there is the so-called summating potential (abbreviated to SP) appearing at different locations in the organ of Corti. SP is an unidirectional electric change superimposed on CM, possibly the result of some sort of additional integration or summation of CM. The function of SP (if any) is not yet known; one can only speculate whether SP in some way interacts with CM in stimulating the nerve endings.

Now that we have identified the hair cells as the microphonic elements of the ear, transducing mechanical vibrations into electric potentials, we must investigate how these cells are built in the ear and how they are reached by the sound waves in the air.

[1] Numbers between brackets refer to the "References", pp. 69–70.

ANATOMICAL LAY-OUT OF THE HUMAN EAR AND THE AUDITORY NERVE

We shall describe this lay-out by following the sound on its way into the human head (fig. 6). [1]

A. THE EXTERNAL AND MIDDLE EAR

The auricle, though of a rudimentary nature, plays a modest part in collecting the sound waves from the air and delivering them to the external meatus, a tube approximately 25 mm long and 6 to 7 mm wide.

At the far end the external meatus is terminated by the tympanic membrane, a thin, tightly stretched sheet resembling a shallow funnel. This cone-shaped structure "collects" the sound pressure in the external meatus and produces a mechanical force necessary to drive the mechanical system of the ear, consisting of the three ossicles, the perilymphic fluid and the cochlear partition.

The most outward of the three auditory ossicles is the hammer, the handle of which is closely attached to the cone of the tympanic membrane. This membrane is kept stretched by a tiny muscle, the tensor tympani, the tendon of which is inserted in the handle of the hammer. By means of a very tight joint the hammer is united with the second ossicle, the anvil. The lower end of the anvil's long process carries a minute spherical nodule, by which the anvil is fastened to the third auditory ossicle, the stirrup or stapes.

The ossicles are contained in an irregularly shaped air chamber, the so-called eardrum or tympanic cavity, one of the walls of which is formed by the tympanic membrane. The air pressure in the

[1] More realistic drawings of the anatomy of the ear are given in figs. 27 and 28.

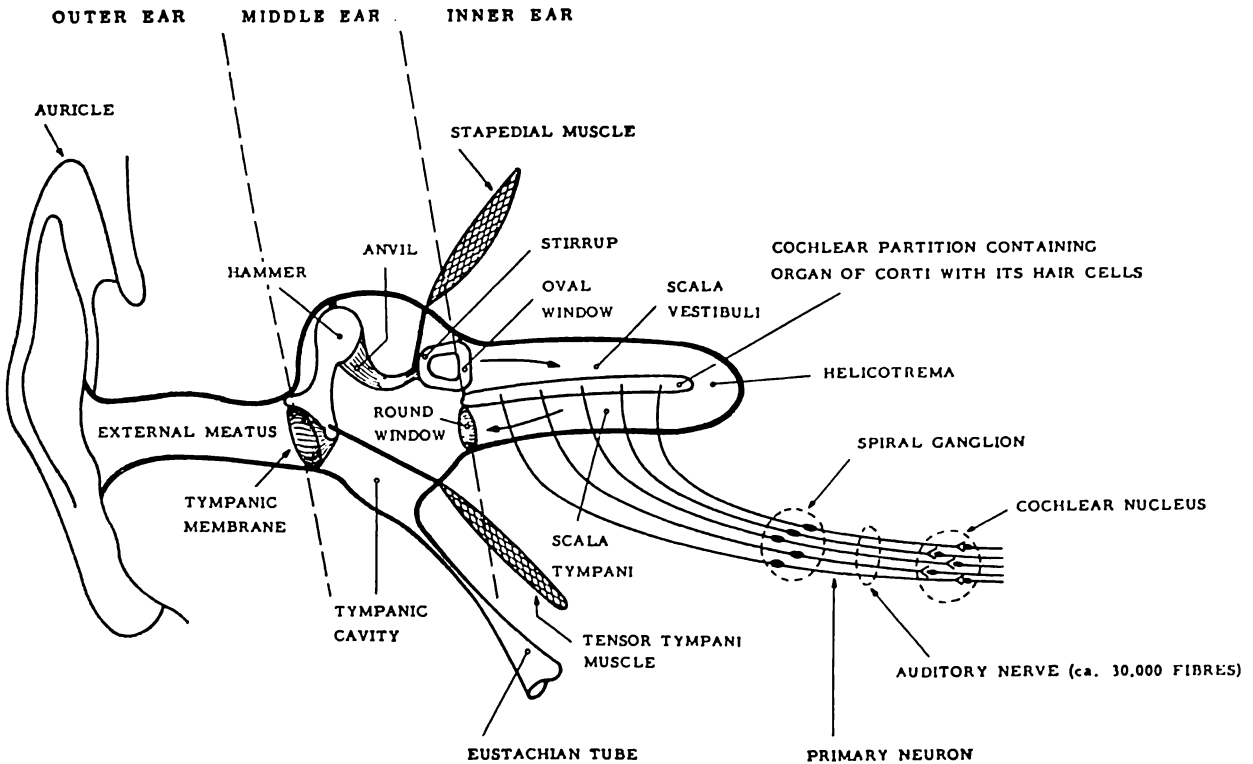


Fig. 6. Schematic representation of the ear and the auditory nerve.

tympanic cavity is regulated by the Eustachian tube that leads to the pharyngeal cavity. Normally this tube is closed but it opens in swallowing, yawning, etc.

The widened end, or footplate, of the stirrup acts as a piston on the fluid content of the bony labyrinth of the inner ear. It is joined to the frame of the so-called oval window by a fibrous tissue with elastic properties. As the fluid, a watery substance called perilymphe, is incompressible, the footplate can only displace the fluid because of the presence of a second window, known as the round window, through which the fluid can "look" into the tympanic cavity. The round window is covered with a thin elastic membrane that can bulge out. Both the oval and the round window prevent the fluid from leaking into the tympanic cavity.

To the stirrup is attached the tendon of the stapedial muscle. This muscle is thought to be able to regulate the axis of rotation of the stirrup. As it only contracts for loud sounds, a protective action as a possible function first comes to the mind.

B. THE INNER EAR AND THE AUDITORY NERVE

The system behind the stapes is called the inner ear. Its fluid content, the oscillating perilymphe, brings into motion the ultimate mechanical link of the ear, to wit the cochlear partition. The inner ear can be considered as a spiralized tube approximately 35 mm long. This spirality, which gives rise to the name cochlea for the inner ear, does not seem to have anything to do with the mode of action of the ear. Therefore in fig. 6 we may depict the cochlea as a stretched structure. The cochlear partition is an elastic structure running along the axis of the cochlea dividing the latter into two parallel parts or scalae. The partition is absent only at the extreme end of the cochlea where the two scalae communicate with each other through an aperture known as the helicotrema. The scala leading to the oval window is called the scala vestibuli, the scala terminating in the round window the scala tympani.

When the stirrup displaces the perilymphe this fluid need not go

all the way around the helicotrema but may push aside the compliant cochlear partition at a region much nearer to the stirrup. The partition in turn pushes the fluid in the scala tympani towards the round window.

As a matter of fact sinusoidal vibrations with a high frequency always prefer the short cut described, leaving the partition and the fluid in the neighbourhood of the helicotrema undisturbed.

The cochlear partition is a very vital part of the ear because it houses the organ of Corti containing the hair cells that transform the mechanical displacement of the partition into an electrical potential known as the cochlear microphonics. As already mentioned, this cochlear microphonic potential is an almost exact electrical replica of the mechanical vibration of the partition. The fibres of the auditory nerve terminate close to the hair cells. Now whereas the hair cells themselves can be regarded as acting practically as linear microphones, the way in which their microphonic potentials stimulate the fibres of the auditory nerve is highly non-linear. Nerve fibres cannot carry potentials with an arbitrary shape, they can only transmit impulses of an "all-or-none" nature, a phenomenon we shall discuss later on.

The nerve endings are bundled together in the auditory nerve which contains some 30,000 fibres. Most of them are interrupted by a ganglion cell in the spiral ganglion, seated in the modiolus, the axis around which the cochlea is coiled up. The cochlear nerve (or auditory nerve) leads to the cochlear nucleus, a switching centre in the medulla oblongata.

Linguistically our chief interest can lie only in the relation between the sound-waves in the air at the entrance of the ear and the nerve impulses in the auditory nerve. We shall, in a functional description of the essential parts of the ear, now study this relationship in greater detail.

FUNCTIONAL DESCRIPTION OF THE EAR AND THE AUDITORY NERVE

A. THE EXTERNAL AND MIDDLE EAR

There is some disagreement on the importance of the auricle for hearing and on the significance of the configuration of the auricle. Perhaps the main function of the external ear is to protect the inner parts of the auditory apparatus from noxious agents of the environment.

A typically electro-acoustic way of describing the function of the tympanic membrane, the ossicles and the fluid load on the stapes is to say that these components together form a so-called low pass filter. In this case this means that the frequencies below 3000 c/s are allowed to pass freely, whereas the frequencies above that limit are attenuated at a rate of at least some 12 dB per octave, corresponding to a factor 4 in amplitude. We may remark in passing that this is the typical behaviour of a highly damped resonator. This high damping is the result of mechanical friction. It is not generally realized that the friction is not located in the ossicles but in the cochlea. The low-pass filter character of the ossicles should not be seen as the result of a careful plan, it is merely the inevitable consequence of a rod-like transmission-system.

As acoustical calculations clearly show, the external ear and the middle ear have the important function of efficiently bridging the gap between two media with entirely different acoustical properties, to wit the acoustically "soft" air and the acoustically "hard" fluid column in the cochlea. It goes without saying that a healthy, unimpaired condition of this transmission-system is of vital importance for the process of hearing in general and for the perception of speech in particular. Nevertheless a detailed study of the middle

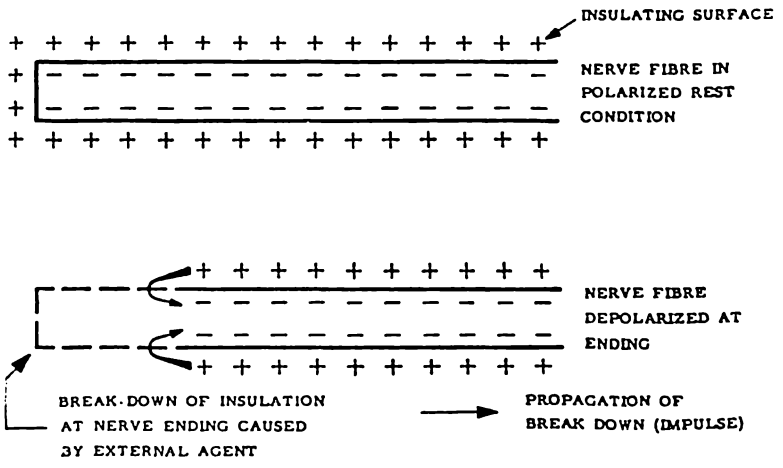


Fig. 7. Schematic representation of a nerve fibre. No anatomical details like the nodes of Ranvier, etc., are given.

ear structures is not worth the linguist's efforts because they only offer feeble resistance to the passing of the phonic information on its way into the human head. The real bottle-neck of the auditory pathway lies further in the cochlear partition. For that reason we shall scrutinize the function of the inner ear and the auditory (or cochlear) nerve more closely in the next paragraph.

B. THE INNER EAR AND THE AUDITORY NERVE

We shall deal with the auditory nerve first because if we know how nerve fibres can be succesfully stimulated we know what kind of action is expected from the hair cells.

Though the almost classic measurements of Tasaki (1, 2) have shed new light on the process of the initiation of nerve impulses, they should not come as a surprise to us in connection with the physiology of a nerve fibre.

A nerve fibre is a very thin cylinder of protoplasm, see fig. 7. Its outside is kept 0.07 volt positive to its inside by the burning of sugar. The positive charge on the outside is separated form the

negative charge on the inside by the surface of the cylinder that in the rest condition of the fibre has excellent insulating properties as long as the voltage between outside and inside is above 0.02 to 0.03 volt positive.

We say that at rest the fibre is polarized. As mentioned above there is a polarisation voltage of 0.07 volt between outside and inside, the outside being positive. As long as the polarisation voltage is between 0.07 and 0.02 to 0.03 volt nothing happens because the surface behaves as an insulator. As soon as the polarisation voltage, due to some external agent, drops below 0.02 to 0.03 volt, or even turns negative after crossing zero, the fibre becomes instable because the insulating surface suddenly breaks down and assumes a very low resistance, almost turning into a dead short circuit. Hence we see that if the voltage is reduced at any small area, current runs in freely, discharging adjacent areas (see arrows in fig. 7) which then short-circuit, draw current and discharge the following areas. Consequently, not unlike a smoke ring, the break-down proceeds along the fibre at a velocity determined by distributed capacity, distributed resistance, and distributed source of energy. The velocities range from about 150 m/sec to 1 m/sec.

When we put a micro-electrode into the core of the fibre somewhere along its course (see fig. 8) and record the voltage the electrode picks up, the passing break-down announces itself as a short, positive impulse, positive because the hitherto negative inner side suddenly jumps to zero voltage.

When the electrode stays in the same place in a given nerve fibre the measured impulses always have the same size. This behaviour gave rise to the notion of the "all-or-none" nature of the nerve impulse, the time of its initiation being its sole characteristic.

It is possible now to make the following predictions as to the mode of action of the inner ear.

As it takes a negative electrical voltage to depolarize a nerve ending, only the negative phase of the microphonic potential produced by a hair cell is able to elicit a nerve impulse.

After having fired an impulse the fibre needs some time for rebuilding its polarization. During that time, the so-called re-

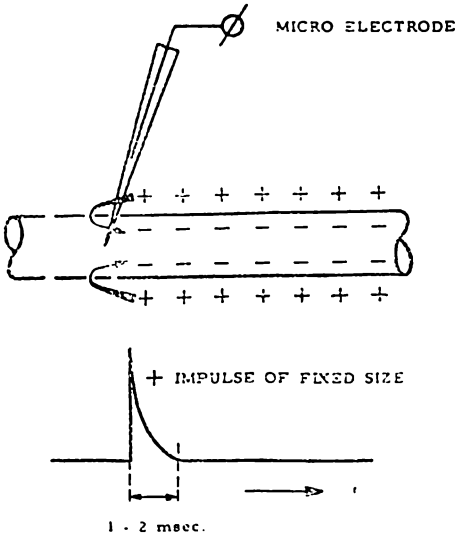


Fig. 8. The all-or-none nature of the nerve impulse.

fractory period, we cannot expect the fibre to be active and to react to the microphonic potential.

When the stimulating microphonic potential is high enough as compared to the threshold of the fibre, we can expect the impulse very early in the negative phase of that potential, that is very soon after the potential shows a zero-crossing from positive to negative. In this respect we speak of a short latency.

When the microphonic potential has to pass through a membrane in order to reach the nerve endings, we can expect a capacitive time-delay so that the fibre will fire close to the negative peak in the microphonic potential, instead of close to the zero-crossing.

The items listed above are supported by experimental findings, among which, in our opinion, those of Tasaki (1, 2) rank first. He measured the impulses in the primary neurons (see fig. 6) of the cochlear nerve in the living guinea-pig, and found they were initiated early in the negative phase of the microphonic potential. It was not possible to predict the appearance of an impulse in a single fibre, but when it came it appeared at the correct moment,

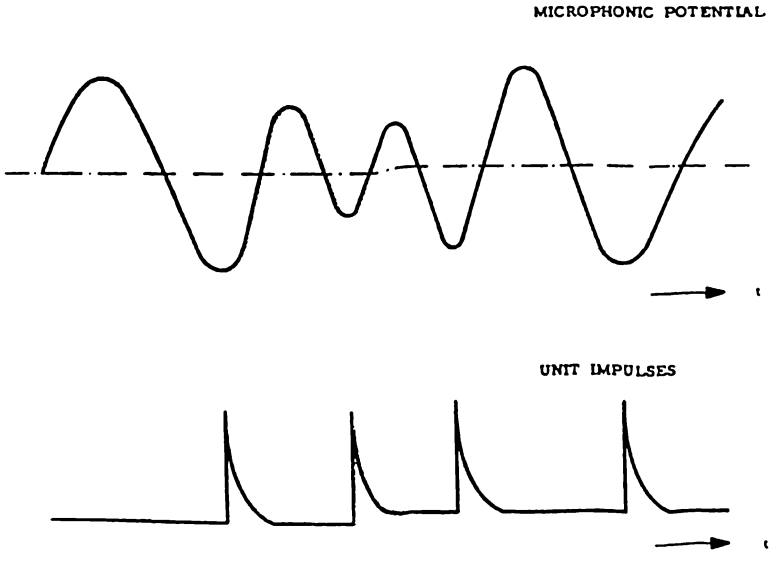


Fig. 9. The mode of action of a mechanism that produces a unit impulse in every negative phase of the curve presented to it.

that is, as already stated, early in the negative phase. The more strongly a fibre is stimulated the more willing it becomes to produce impulses at a higher rate than once in a while, until at last the fibre is saturated and delivers an impulse in every negative phase, provided the refractory period permits it to do so. When several fibres are co-operating there is a high probability that one fibre will transmit an impulse during the refractory period of its inactive neighbour, so that a bundle of fibres is able to produce a reaction in every negative phase, even when the constituent fibres are not yet saturated. Fig. 9 is the graphical description of the expected behaviour of a bundle outlined above (6). It enables us to see at a glance how the bundle will react to the microphonic potentials presented to it.

It is interesting to notice that all theories of hearing can be reduced to the way in which the just mentioned elementary bundles are supposed to cooperate. The bundles differ in that they are stimulated by different microphonic potentials.

When, in general, a certain sound curve is fed into the ear, the time course of the microphonic potential depends on the distance from the stirrup at which it is measured.

In the so-called basal turn, the region of the cochlear partition near the stirrup, the time course of the microphonic potential shows more high-frequency detail than in the apical turn, the region near the helicotrema. In other words: the cochlear partition is able to change the shape of the time course of the microphonic potential. Its activity can be described as *curve-shaping*. This activity has a mechanical background which we shall discuss in more detail in a separate chapter.

Before, however, we shall briefly mention how the bundles come into contact with the hair cells.

In fig. 10 we see how the cochlear partition separates the scala vestibuli and the scala tympani. It is bounded by Reissner's membrane, the stria vascularis and the basilar membrane. On the latter membrane rests the organ of Corti, containing the hair cells. The cochlear partition, also called the scala media, is filled with an oily fluid, the endolymphe furnished by the stria vascularis. As there are no blood-vessels in the organ of Corti the endolymphe has the task of feeding the structures in the scala media.

As already mentioned, when the stirrup moves to and fro it displaces the perilymphe in the scalae. As a result the cochlear partition moves up and down, though not necessarily over its entire length. Neither should we expect that the excursions of all moving parts of the partition display the same time course. Though the basilar membrane moves up and down, von Békésy has shown (10), by artificially moving the basilar membrane, that the hair cells are most sensitive to radial displacements. Therefore we may expect that the hair cells derive a radial displacement from the vertical movement of the cochlear partition.

Figure 11 presents a more detailed sketch of the organ of Corti. After shedding their myelin sheaths the fibers of the auditory nerve enter the cochlear partition through numerous holes in a bony ledge that follows the organ of Corti along its entire length. As the organ of Corti moves up and down the fibers act as flexible

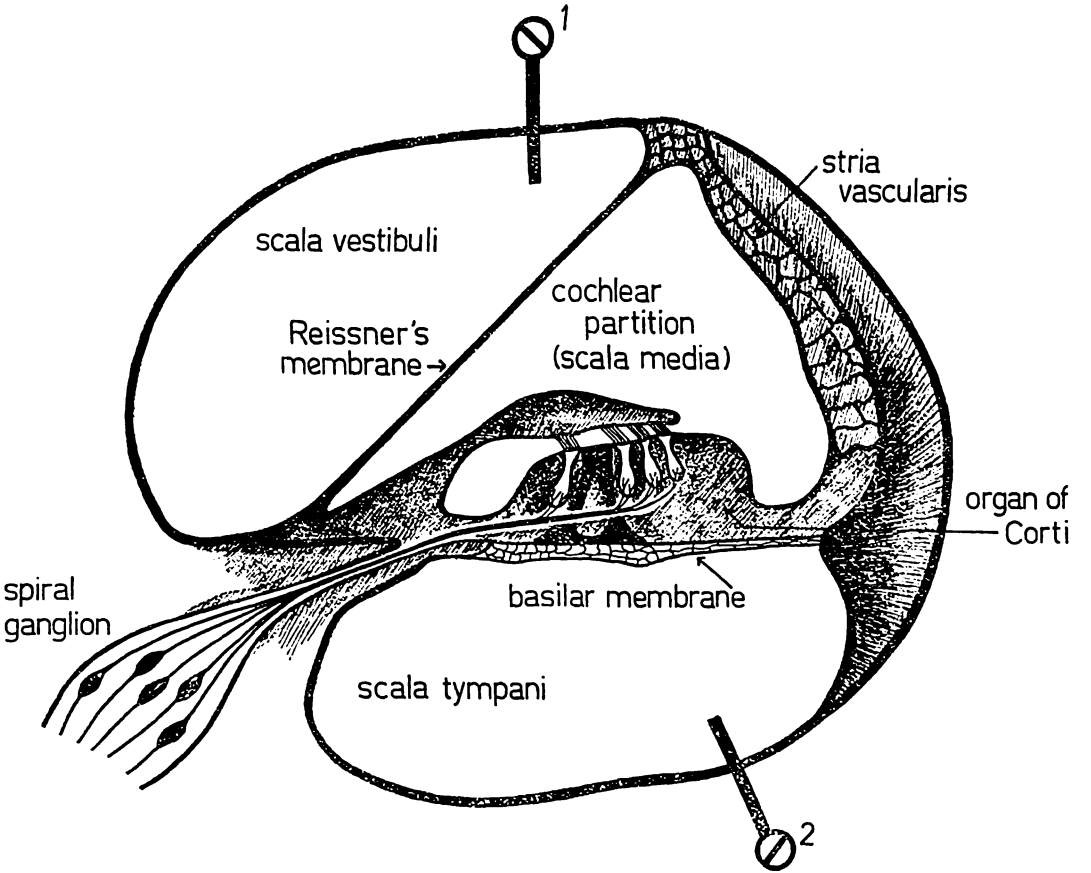


Fig. 10. Cross-section perpendicular to the axis of the cochlea. 1 and 2: electrodes between which the cochlear potential can be measured.

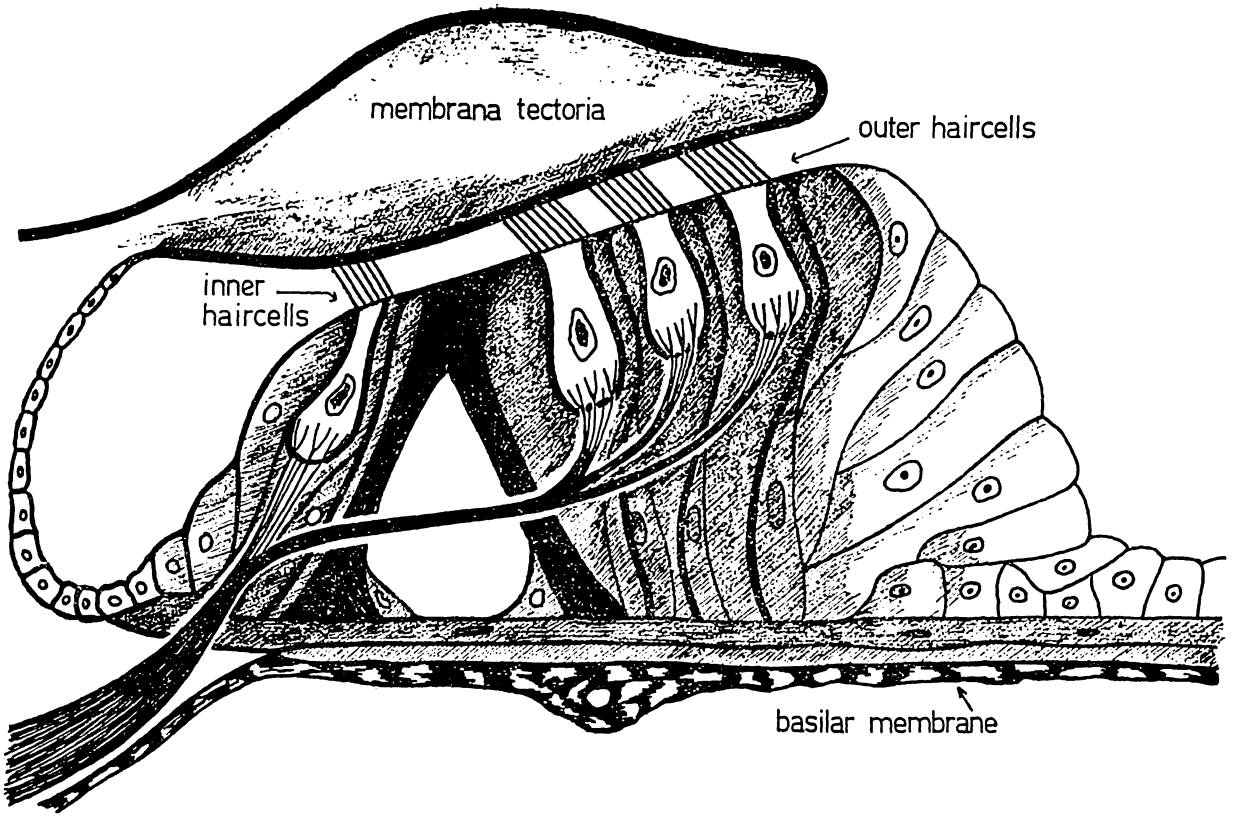


Fig. 11. Detail of the cochlear partition showing the organ of Corti.

wires connecting the moving hair cells to the nervous system.

There are two groups of hair cells. The so-called outer hair cells are located at the site where the organ of Corti makes its biggest excursions. Therefore they are more sensitive than the inner hair cells which more or less rest on the bony legde. Whether the distinction between the inner and outer hair cells is necessary to the perception of speech is dubious because that distinction does not exist in the ear of a parrot, a bird capable of imitating speech sounds in a very satisfactory way, and thereby proving that its simple ear is adequate for the purpose.

Electron microscope studies have shown (11) that there are two types of nerve endings on the hair cells, schematically shown in

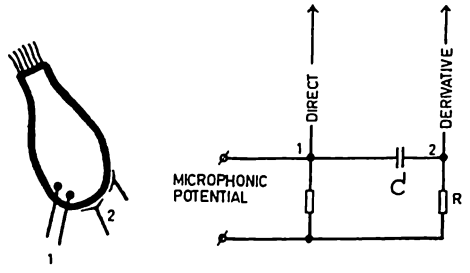


Fig. 12. Two types of nerve endings on the hair cells.
Also the electrical circuit is given.

fig. 12. Some nerves end as simple buttons (type 1) while others form clusters that cover an area of 4 to 5 μ applied closely to but not penetrating the hair cell membrane (type 2). We suppose here that a nerve ending as a button has direct access to the microphonic potential whereas the clusters reach that potential via a condenser, formed by the hair cell membrane and the membrane of the nerve ending. When, in general, a voltage has to pass through a condenser its wave-shape is altered. An example is given in fig. 13.

It can be shown mathematically that the new shape depends on the choice of the condenser C and the resistance R. As a matter of fact, for a suitable value of the product RC, also called the time-constant, the new shape practically represents the slope of the original curve. Mathematically speaking, the new curve is the

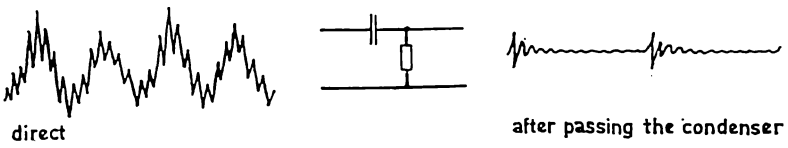


Fig. 13. Wave shaping as effected by a condenser in series with a resistance (electrical wave shaping).

first derivative of the original curve. The readers, however, who want to skip mathematics, can simply look at fig. 13 in order to see the influence of the condenser.

We draw the attention of the reader to the fact that the type of wave shaping, just discussed, is peculiar to the hair cells. It has no connection with the mechanical behaviour of the cochlear partition.

THE POSSIBILITIES OF CURVE SHAPING IN THE COCHLEA

The first to suggest that the cochlea performs a mechanical wave analysis was von Helmholtz, who assumed the radial fibres of the basilar membrane to act as stretched strings, each string resonating to the frequency to which it was tuned.

Today this so-called piano theory is no longer tenable, especially after von Békésy showed, by touching the basilar membrane with a fine hair, that there is no tension in that structure. So the idea of sharply tuned stretched fibres has to be abandoned. The same goes for the conception that went with the stretched strings, i.e., again according to von Helmholtz, the cochlea is the material realization of a mathematical method of decomposing sound curves into sinusoidal components known as Fourier-analysis of frequency-analysis. But though the stretched strings are nowadays discarded, most investigators desperately cling to the idea that no matter what happens the ear performs a Fourier-analysis. In this chapter, however, we shall show, that the old question whether the cochlea is able to perform a Fourier-analysis or not can be re-drafted as follows: *does the ear need the complete length of the 33 mm long organ of Corti in order to deal with speech waves?*

Let us first explain why the mechanical behaviour of the lenghty cochlear partition is not uniform in all points.

In his thesis Wansdronk (12) has recently reviewed the existing theories of the dynamics of the cochlea (Fletcher (13), Zwislocki (14), Peterson and Bogert (15), Wansdronk (12)). All these theories are based on the same principle, differing mainly in the approximations made in the calculations.

Only a small layer of liquid at both sides of the cochlear partition plays a part in its motion. In other words, the cochlea contains an excess of perilymphe outside the thin layers. Mechanically speaking

the cochlear partition may be considered as being composed of more or less independent elements coupled to each other by the liquid in those thin layers. The stiffness coupling of neighbouring elements can simply be neglected. The stiffness of the cochlear partition is not constant but decreases exponentially from stirrup to helicotrema. Its mass is either neglected or considered as a constant. Wansdronk did not solve the mathematical motion-equation he had derived and compared with those of other investigators.

Equations of that type can be solved only very laboriously by numerical methods. Instead he built an electrical model based on the mechanical constants of the cochlear partition, with thin liquid layers at each side of it. Needless to say this model represents an approximation only as its limited (though big) number of components cannot guarantee an exact imitation.

Nevertheless one gets a fair impression of how the points of the cochlea change the shape of the sound curve which is furnished by the middle ear. As the model has 100 electric taps corresponding to 100 different points on the cochlear partition, one can study by means of the cathoderay-oscilloscope how differently located points change that curve in their own way. Fig. 14 shows schematically how the model is utilized. By working with models one need not sacrifice so many laboratory animals in order to come to conclusions.

When one describes the mode of action of the cochlea in terms of its wave shaping properties, one has no longer any use for sinusoidal waves. When a sinusoidal vibration with, for instance, a low frequency is presented to the stapes (= stirrup) the whole cochlear partition between stapes and helicotrema is vibrating, and every point of the partition will execute a sinusoidal vibration. When the frequency is raised the vibrating part is seen to retreat towards the stapes so that a region near the helicotrema remains undisturbed (4).

The actually vibrating part, however, executes sinusoidal vibrations. In other words: wave shaping does not come into action in this way: a sine wave at the stapes remains a sine wave at any vibrating point of the cochlear partition. Nevertheless, testing a

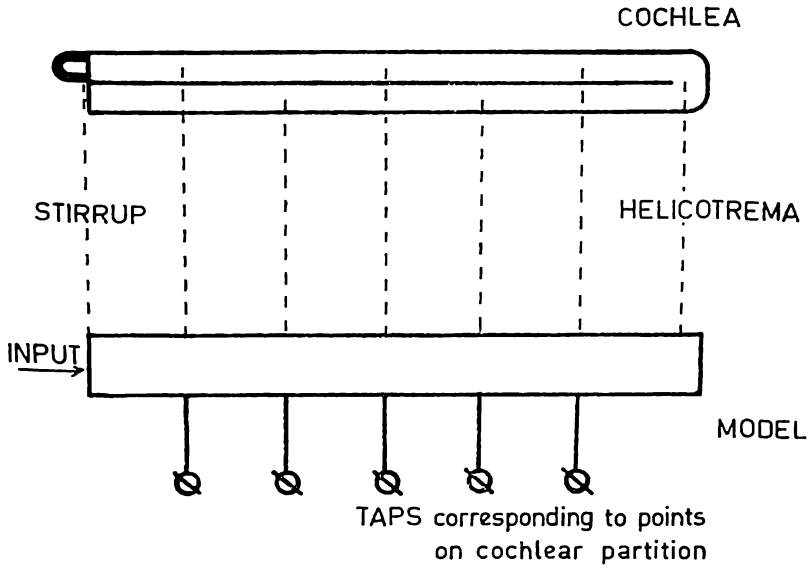


Fig. 14. Using an electrical model of the cochlea for the study of the curve shaping properties of the partition.

model of the cochlea by studying its reaction to pure tones, as sinusoidal waves are often called, can be useful, especially when one wants to check whether the model in question agrees with the differential equations pertaining to the real ear. Also for diagnostic purposes pure tones can be applied in audiometry in order to get an impression of the location of defective parts on the cochlear partition.

These experiments, however, need not be directly related to the *mechanism* of aural identification of speech waves. One can prove, for instance, *that* a certain part of the cochlea is vital for recognizing certain speech sounds, not *why* it is vital, which makes all the difference.

Wansdronk showed that when vowel sounds were presented to his model the sound curves recorded at the different taps were certainly not sinusoidal but had the character of damped oscillations instead. In other words: the mode of action of the cochlea is not analogous

to a narrow band Fourier-analysis. The curves from taps near the helicotrema show the closest resemblance to damped oscillations because the damped oscillations corresponding to the second and higher formants, have been effectively removed there so that only the low pitched first formant remains. Points nearer to the stapes show a mixture of damped oscillations corresponding to all formants. In that region, called the basal turn of the cochlea, there is no separation of a mechanical nature betweeh the formants. This statement is in harmony with a quotation from a paper of Tasaki (2), who measured the nervous activity in nerve fibres leading to the different regions of the cochlear partition: "The cochlea has long been considered as a kind of wave analyzer which is capable os separating a compound sound wave into its components. Thif notion, however, is only partly true. In the basal turn a mixture of high and low tones causes a mechanical vibration as such, namely, without its being resolved into its components, and excites the nerve endings in the form of the applied mixed wave. Separation between the components occurs only as the mechanical wave (caused by the mixed tones) travels along the cochlear partition upwards and the higher frequency component decays more rapidly than the lower one as they travel. The application of Fourier analysis to a complex sound wave and interpretation of the total physiological effect as a sum of the effects of those constituent pure tones is dangerous and in most cases erroneous.

Wave analysis is of course essential for any theoretical and practical treatments of dynamical problems in the cochlea; but for the consideration of the process of initiation of nerve impulses (which are all-or-none in nature), particularly in the basal turn where any mixture of tones can act without being separated into its physical components is undoubtedly worthless."

Though the curves at the taps of Wansdronk's model were not sinusoidal he measured their peak values.

By means of a special cathoderay-oscilloscope he could depict those peak values (called amplitudes for the sake of convenience) as a function of the place of the tap. He called those graphs the cochlear spectrum, a rather misleading term because what is shown

is certainly not the result of a Fourier-analysis but we shall not stress that point here.

Figures 15 and 16 show some of his results for several different Dutch vowels.

The sound curves presented to the model are depicted in the left-hand columns whereas the right-hand columns present the corresponding cochlear spectra. In the spectra the stapes is to the left.

We must be very careful in evaluating Wansdronk's cochlear spectra. At first sight they are rather attractive as they show the formants as relative maxima in the envelope.

Every type of vowel has its maxima at different positions on the cochlear partition. The question, however, is whether the hair cells are able to send to the brain detailed information on the position of these maxima. For that purpose the following mechanisms are needed.

In the first place, a mechanism that measures the amplitudes of all points of the partition at a certain moment and stores them in a memory.

Second, a scanning mechanism that scans the memorized amplitudes and indicates the maxima and their positions.

Both mechanisms must be able to perform their task in a very short time.

It is highly improbable that the mechanisms mentioned can be found in the ear. But, nevertheless, Wansdronk's model proves the presence of a mechanical selective system, though its resolving power is rather limited and it needs an elaborate nervous network to evaluate its results.

The presence of a long cochlea is essential for the existence of the cochlear spectrum. It is at this point that we begin to doubt its function.

It is known (31, 32) that birds have a rudimentary cochlea. The parrot has a cochlea of less than 3 mm so that no mechanical cochlear spectra are possible.

Also, in the ear of the bird there is no clear distinction between the inner- and outer hair cells. There is no pinna and there is only one ossicle, the so-called columnella.

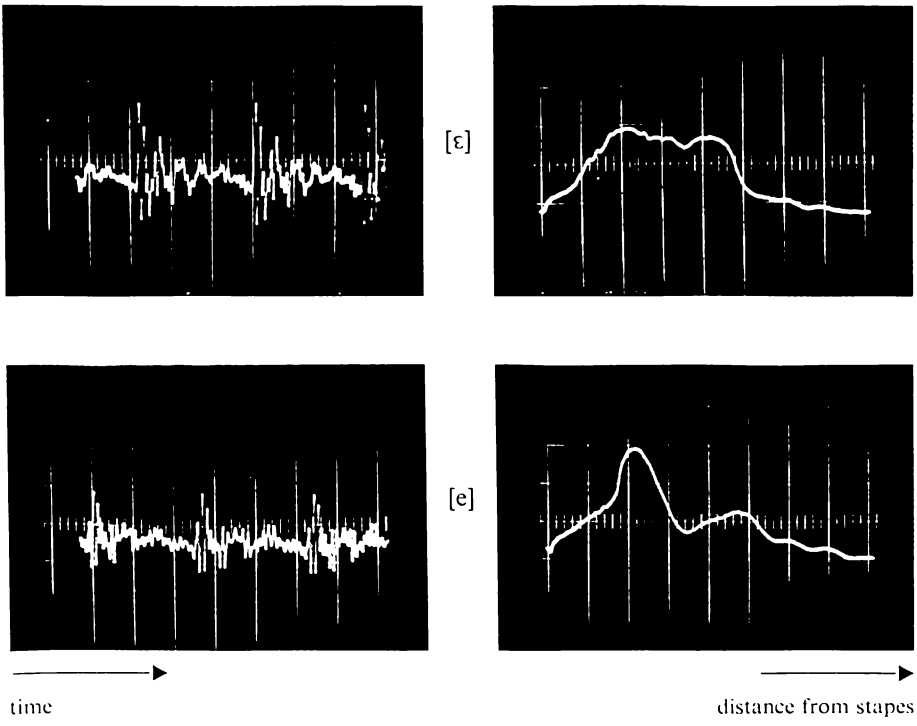


Fig. 15. Oscillograms (sound curves) and cochlear spectra (as defined by Wansdronk) of the Dutch vowels [a], [ɛ] and [e]. The spectra are in the right-hand column.

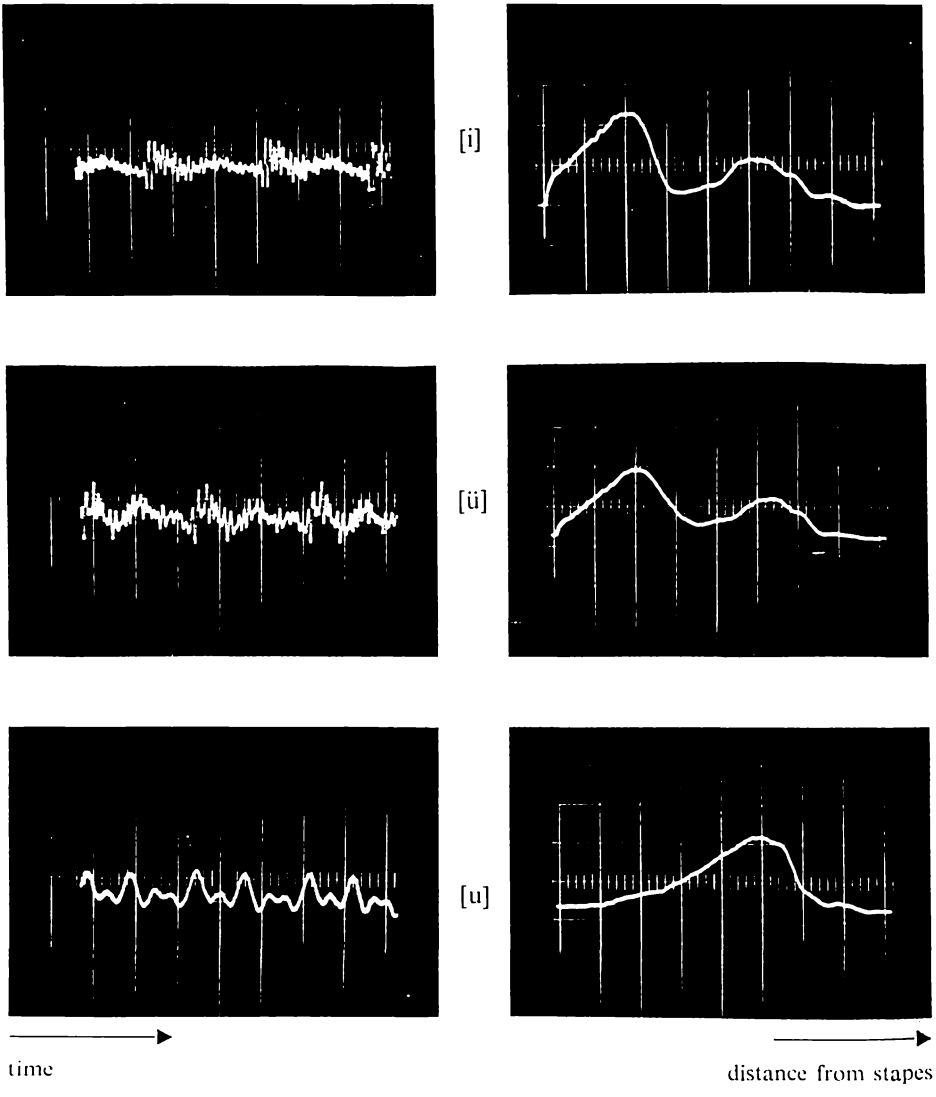


Fig. 16. Oscillograms (sound curves) and cochlear spectra (as defined by Wansdronk) of the Dutch vowels [i], [ü] and [u]. The spectra are in the right-hand column.

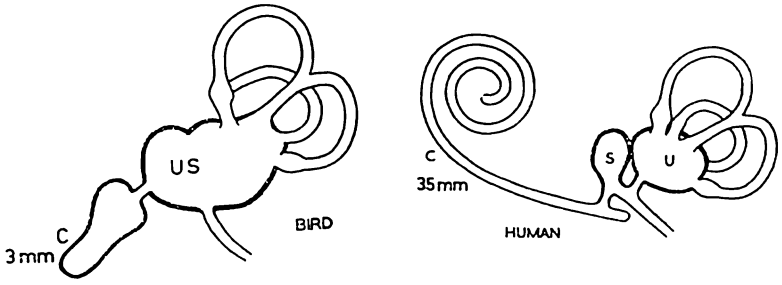


Fig. 17. Comparison between the human cochlea and the cochlea of the bird. U = utricle, S = saccule, C = cochlea.

Nevertheless the parrots can produce all human speech sounds in a satisfactory way which proves that its ear has effectively dealt with those sounds in spite of its rudimentary nature.[1]

Furthermore, as measured with pure tones, the frequency range of the ear of the bird (40–20000 Hz) is not fundamentally inferior to that of the human ear (20–20000 Hz). The same is the case with frequency discrimination (3).

In fig. 17, a comparison is made between the human cochlea and that of the bird. The difference in length is striking: 35 mm as opposed to 3 mm. The rudimentary cochlea of the bird can be considered as a human cochlea which has been cut down to a length of 3 mm. Consequently the possibility of housing a cochlear spectrum of sounds is nonexistant. It is interesting to project the limited cochlea of the bird on that of the human ear, as can be seen in fig. 17a. We indicated that projection as the "parrot-zone". As that zone, in the ear of the bird, is amply sufficient for speech perception, it is not too far fetched to suppose that in the human ear too the parrot-zone is responsible for speech perception.

It seems that the development of the long cochlea of the mammals is far in excess of the needs of speech perception. The long cochlea must serve another purpose. We are inclined to speculate it has

[1] The parrot even succeeds in imitating the personal characteristics of a particular speaker. The current, but apparently wrong, ideas on that subject are that a very complicated mechanism of hearing is needed for that purpose.

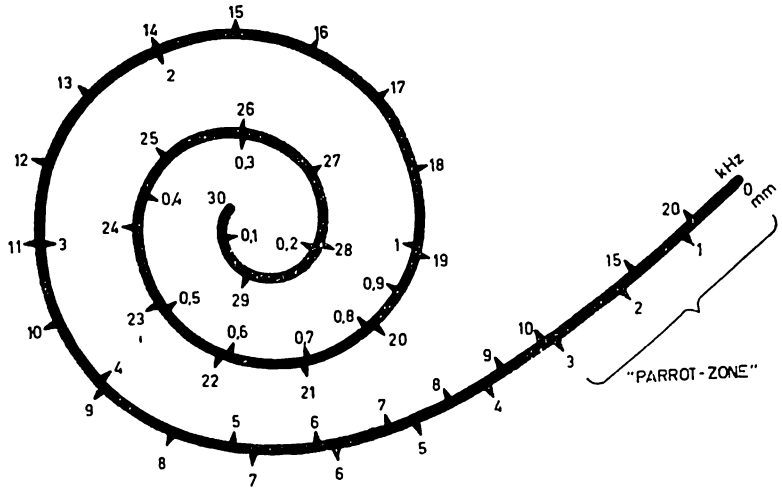


Fig. 17a. The "parrot-zone".

something to do with stereophonics, the art of locating sound sources.

In the past (5, 6) the author was engaged in constructing simple electrical models that could simulate the activity of the organ of hearing in detecting speech sounds. Before 1961, however, we never went to the length of merely imitating the parrot-zone: we still clung to the complete human cochlea.

In 1959 Uhlenbeck and Mol (6) showed that it is sufficient to take into consideration the sound curves at only *two fixed* taps of a cochlear model provided the positions of these taps are well-chosen. A tap corresponding to the middle of the cochlear partition will produce a damped oscillation representing the second formant of a vowel whereas a tap corresponding to a point near the helicotrema will show a damped oscillation representing the first formant. When the damped oscillations are fed into an impulse producing mechanism described in fig. 9 the periods of the formants will appear as time intervals between the impulses.

Wansdronk did not measure at two fixed taps of his model. He selected the taps where, for a certain vowel, the displacement of the cochlear partition showed a relative maximum. He scanned

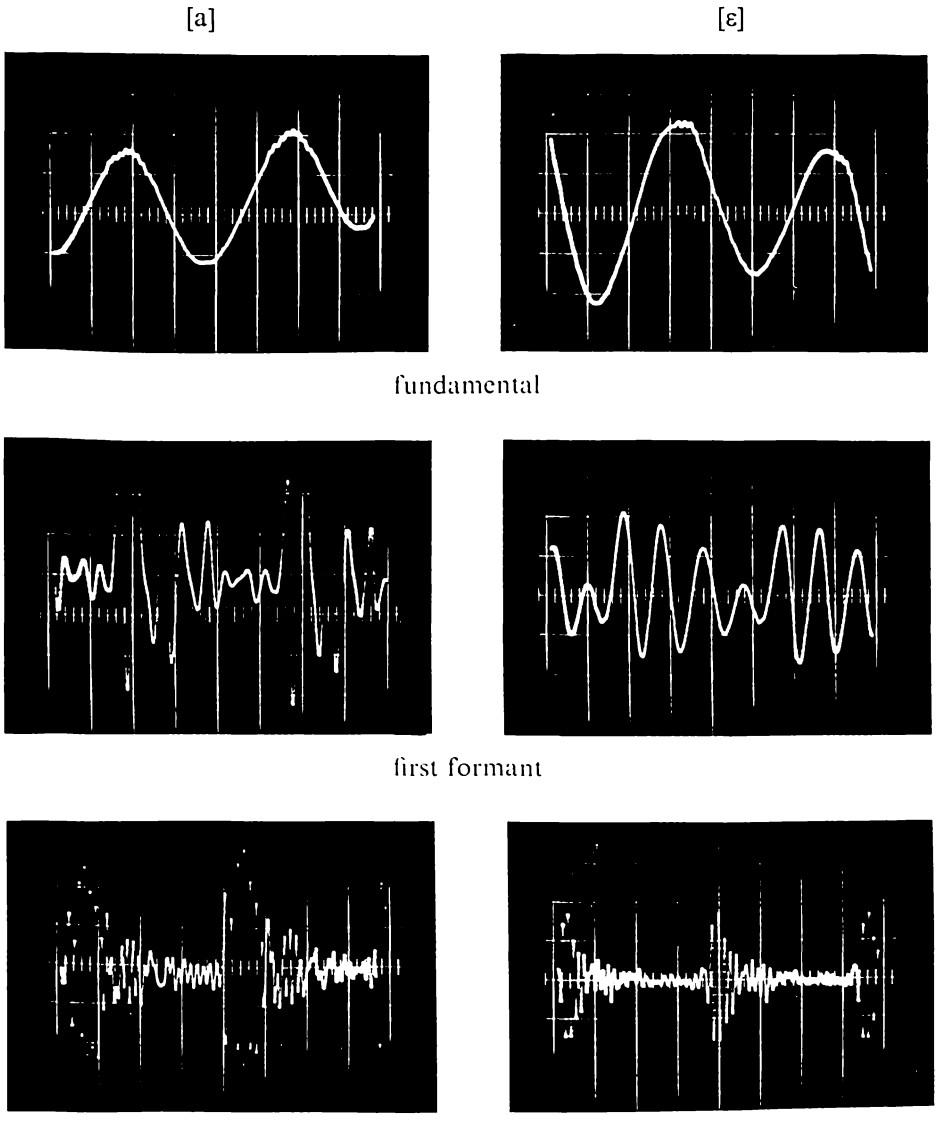


Figure 18. Signals found at different taps for the vowels [a] and [ɛ]. (After Wansdronk.)

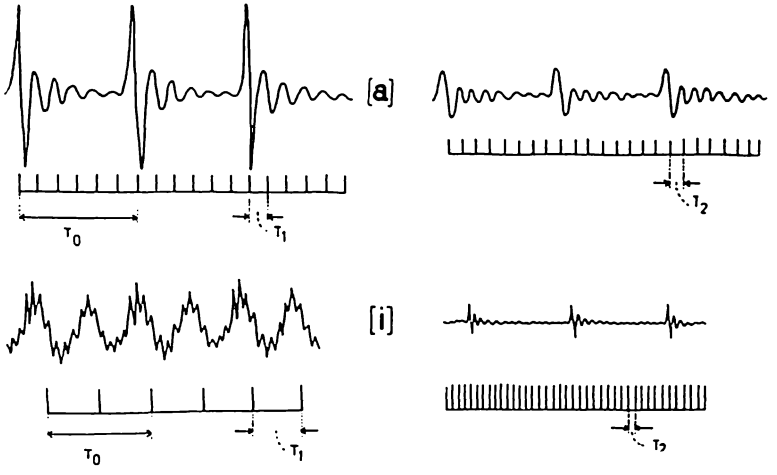


Fig. 19. Performance of a parrot-zone model of the cochlea for the vowels [a] and [i]. The left-hand column corresponds with the first formant whereas the right-hand column refers to the second formant. The periods of the formants are indicated by the time intervals between the impulses.

his cochlear spectra for maxima, see fig. 15 and 16, and defined these maxima as the two formants. He had the advantage of automatically getting only two formants in his spectra. His spectra do not offer the confusion produced by the usual spectrogram of a spectrograph where many bars appear from which one has to select the first two formants.

Wansdronk also recorded the oscillograms at several taps but he did not actually use them like Uhlenbeck and Mol in their two-tap model. Some examples are given in fig. 18 where the vowels [a] and [ɛ] are treated.

However simple a two-tap model may be it is still based on a full-length cochlear partition. As such it cannot explain the verbal behaviour of the parrot with its rudimentary organ of Corti.

In a parrot-zone model no use can be made of the mechanical properties of the cochlear partition because there is no mechanical wave shaping active in this part of the basal turn.

Instead we must make use of the method of electrical wave shaping depicted in the figures 12 and 13. This method is supposed

to be used by the hair cells themselves. The mechanism described in fig. 9 is nothing but an artificial bundle of nerve fibres in which there is an impulse for every negative phase of the stimulating potential.

Figure 19 elucidates the mode of action of the model.

When the cochlear potential directly stimulates an artificial bundle there appear impulses corresponding to the first formant. When the derivative of the cochlear potential stimulates an artificial bundle there appear impulses corresponding to the second formant. There are only two formants because there are only two known types of nerve endings.

As can be seen in fig. 19 the pitch of the vowels is represented by the fundamental period T_o.

This period, however, is not present as the time interval between two adjacent impulses.

Because we do not wish to saddle the reader with technical details we confine ourselves to stating that, by combining artificial bundles, a new mechanism can be created that produces impulses spaced at time intervals of T_o, the fundamental period.

In other words, the combined efforts of very simple mechanisms in the parrot-zone can deal with the two formants and the fundamental pitch of the vowels.

SUMMARY AND CONCLUSIONS

1. The most important elements of the organ of hearing are certainly the hair cells situated in the vibrating cochlear partition. They transform their mechanical movement into an electric potential (also called microphonic potential or cochlear microphonics, abbreviated as CM) that, either directly or indirectly, stimulates the nerve endings. The hair cells behave like amplifiers.

2. There is strong evidence that in the best conditions in each negative phase of the microphonic potential at a certain point of the partition a nerve impulse is generated in a small bundle of fibres leading to that point.

3. The importance of the highly developed 35 mm long mammalian cochlea for speech perception is extremely doubtful because the parrot with its less than 3 mm long rudimentary cochlea is able to imitate all speech sounds in a satisfactory manner.

4. Models show, however, that in the long mammalian cochlea there is a mechanical possibility of roughly separating two formants, but the nerve mechanism needed for the evaluation of that selective mechanical behaviour of the cochlear partition is so complicated and time-consuming that it is reasonable to suppose that in the human ear too speech sounds are dealt with by a limited zone in the basal turn labelled by us "parrot-zone".

5. In the parrot-zone there is no mechanical separation of the formants. Nevertheless the two formants can be extracted as can be shown in electrical models, based on the principle that at the hair cells the microphonic potential is split up into two components

that are separately fed to the nerve-endings. One component is supposed to be the microphonic potential itself. The second component is thought to be its derivative, that is the rate of change of the cochlear potential.

6. Though it is possible, in a model, to extract and measure the second formant by means of the method of the first derivative it is not sure, however, that the hair cells employ exactly the same method. When listening to artificial vowels consisting of only two formants we can sometimes readily detect changes in the second formant which proves that the hair cells can in any case in some way deal with it, probably applying a sort of wave-shaping akin to a derivative. We are inclined to believe that an additional mechanism, perhaps betraying itself as the already mentioned summating potential is active in cases where the second formant is very weak as compared with the first formant.

Further research is needed for clearifying that point. We would not even be surprised if future research showed that the second formant cannot be determined by the ear in all cases with certainty.

7. For vowels the components show the shape of damped oscillations the periodicity of which can be measured by the model of a small nerve bundle. The periodicity comes to the fore as the time interval between two adjacent nerve impulses.

8. In a short cochlea the fundamental pitch of a vowel can certainly not be indicated as the presence of the Fourier-component with the lowest frequency. The parrot-zone in the human ear can probably deal with fundamental pitch because it can measure the time interval between the successive air puffs emitted by the vocal tract. The parrot is able to imitate voice intonation with great precision.

9. Though the part the ear plays in the recognition of vowels is more or less clear, the way in which the consonants are identified is still rather obscure. As in normal speech the articulatory movements accompanying the consonants often influence the

positions of the formants of the adjacent vowels, the transitions (shifts) of the formant positions of a vowel might be helpful, though not the only, cues for the identification of adjacent consonants. In this view (part of) the consonants are parasites that live at the expense of the vowels. In consonant clusters, however, the consonants must be self-supporting. In Czech there even exists the cluster [k] [p] [s] [t] [r]. A parrot-zone model can in any case detect formant shifts. For the identification of other consonantal cues cooperation between nerve bundles leading to different hair cells seems indispensable. We shall not, however, in this monograph speculate too much on these possibilities before further research has given us a safer foundation.

AURAL STIMULI AND THEIR INTERPRETATION*

The title of this chapter draws the attention to the fact that in speech aural stimuli call for interpretation. The comparatively recent development of electro-acoustics has provided the investigators with the technical means to prove that the aural stimuli do not directly label the phonemes of the words understood by the listener. The mechanism of recognizing spoken words and sentences is no doubt voice-operated but it may not be regarded as the acoustic counterpart of an electrical teleprinter which prints letters that are unambiguously labelled by the electric signals it receives from a transmission line. When, in a teleprinter system, a key on the keyboard of the sending typewriter is pressed a normalized electrical signal corresponding to that key is generated and sent along a line to the receiving machine to print one special letter, to wit the letter corresponding to the signal and the pressed key on the sending machine. The receiving teleprinter has no choice, it acts as a slave and no interpretation is expected.

Engineers who try to construct so-called voice-operated typewriters and speech recognition systems experience that the process of speech recognition is not analogous to the teleprinter system. In spite of the hopeful announcements in the popular press a voice-operated typewriter can perhaps be made to react reasonably to its master's voice but it fails to cope with the variety of voices with which a human listener has no difficulties. The machine needs a brain.

* The following chapter is based on a paper read by the author at the Fourth International Congress of Phonetic Sciences (Helsinki, 1961). It can be read independently from the other chapters. The reader will no doubt notice that in this chapter several problems are briefly summarized which have already been dealt with in more detail in the foregoing chapters. This chapter must be regarded as a first attempt to pave the way for a better understanding of the process of speech recognition.

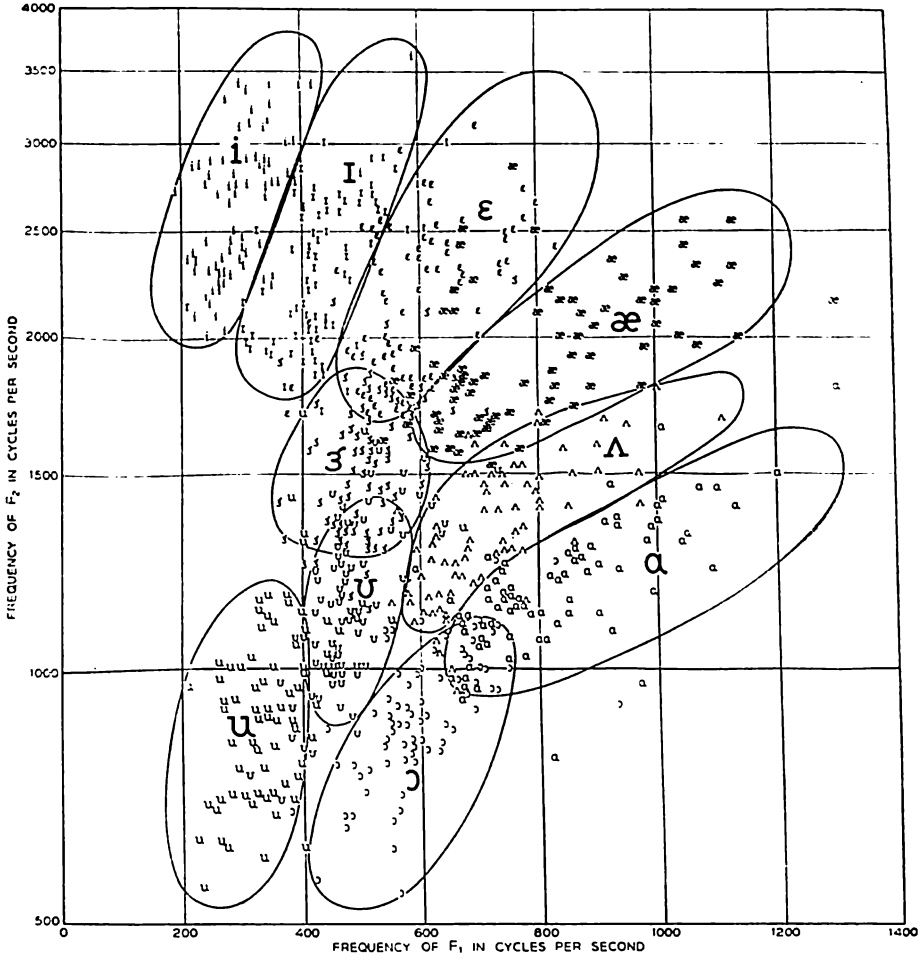


Fig. 20. Frequency of second formant versus frequency of first formant for 10 vowels by 76 speakers (Peterson and Barney).

The following examples, selected from recent publications, illustrate the absence of the much hoped for analogy with the teleprinter.

In the almost classic measurements of Peterson and Barney (18) on vowels 76 talkers pronounced 1520 isolated words containing

10 American vowel types. The first two formants of every vowel were measured and plotted in the well-known way of plotting F_2 horizontally and F_1 vertically, see figure 20. They found 10 targets around which the realisations spread, but although the talkers pronounced isolated words there was considerable overlap between the targets. When the 1520 words, recorded on magnetic tape, were presented to a jury of 70 listeners only about one half of the words were unanimously correctly recognized.

By carefully selecting a new jury this fraction could be raised to about two thirds. These experiments show that even for isolated monosyllables there was uniformity neither among the talkers nor among the listeners.

Ladefoged and Broadbent (19), in 1956, showed that the recognition of a synthesized word could be influenced by the phonetic character of a likewise synthesized, introductory sentence. This experiment confronts us with the amazing ability of a listener to adapt himself to the articulatory habits of a speaker in a very short time.

In 1957, Blom and Mol plotted the formants of 12 Dutch vowel types appearing in a simple sentence freely pronounced by a large number of talkers (7). For this running speech they did not find 12 targets but only 2 targets instead, showing that the talkers, when left to themselves, reduced distinction to a bare minimum. One target contains the vowels [i], [ɪ], [ü], [ə], [ö], [e] and [ɛ]. The other target embraces the vowels [a], [ɑ], [o], [ɔ], and [u], see fig. 21.

The lesson of all those investigations is that the phonic data the ear extracts from the sound waves do not form the only source on which the listener bases his identifications.

We can also derive the warning from the just mentioned and similar investigations that it is dangerous to draw conclusions on the mode of action *of the ear itself* from hearing tests in which speech is used as a stimulus. There is always the problem of separating the passive action of the ear from the interpretative activity of the brain. Let us try to start from scratch in order to get a clear picture of the problems involved.

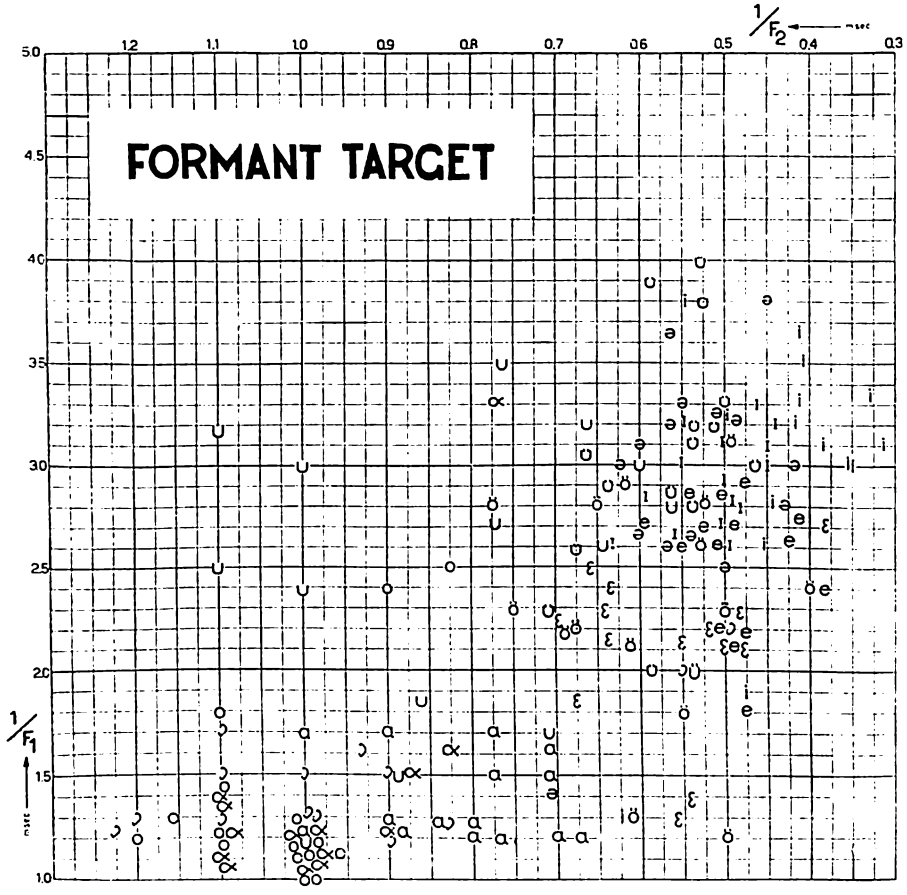


Fig. 21. Period of second formant versus frequency of first formant for 12 Dutch vowels appearing in a simple, freely pronounced sentence. The data of this pilot investigation refer to 15 talkers taken at random from a group of 100 talkers. The numbers on the axes represent periods, in harmony with the fact that the formants were indeed measured as periods and not as frequencies.

The sound waves produced by the articulators and, in fact, all sound waves, are transformed by the ear into nervous quantities suitable for processing and interpretation by the nervous system. This transformation is performed in the organ of Corti by a very big number, more than say 23,500, of surprisingly simple sensory

elements that can only say yes or no when driven by the excursions of the partition containing them. Galambos (3), in his excellent review of 1954, also drew the attention to the simplicity of the sensory elements. Traditional theory attributes an important selective power to the mechanical and hydrodynamical action of the cochlear partition and perilymphe. It speaks of a spatial spread along that 35 mm long partition. But the parrots and other speaking birds, many of which display a remarkable capacity of producing fully recognizable speech, have only a rudimentary cochlea with a partition of merely 3 mm. Furthermore they do not have a clear distinction between external and internal hair cells like the mammals display. The fact that some birds can speak throws doubt on the necessity of a long cochlea for the precessing of speech and indicates that the simple sensory elements themselves can perform some type of analysis.

Since the outstanding work of Tasaki (1) we know that a sensory element can *at best* induce a nerve impulse in a single fibre of the auditory nerve every time the partition containing the organ of Corti passes its position of equilibrium in the direction of the oval window. In fact this is the well-known technique of indicating zero-crossings in a sound wave.

Indeed in Lingua, Uhlenbeck and Mol (6) described how this simple mechanism allows formant extraction. We did not, however, present a complete and definite description of the mode of action of the ear but merely showed the tremendous achievements of simple mechanisms in order to pave the way to a better understanding of the mechanism of hearing.

In a recent issue of *Nature* (20), Mr. Broadbent, discussing the perception of pitch, says: "One may argue that the pitch of a sound will depend not so much on the precise receptors stimulated (as traditional theory holds) as on some features of the message travelling up the nerve fibres."

"A likely candidate," Mr. Broadbent suggests, "is the frequency of impulses in the auditory nerve."

It is exactly this principle which is incorporated in the mechanisms described by Uhlenbeck and Mol.

The only sage way to discover the role the organ of hearing plays in the phonemic classification of sound waves is to perform experiments on real ears with the aim of measuring the nervous activity in the fibres of the auditory nerve for a given sound wave.

A given sound wave is defined here as a sound wave of which the time-function is exactly known.

Up to now most experiments have been performed with pure tones which are indeed time-functions precisely known. As the production of a nerve impulse by a sensory element is highly non-linear, however, the reaction of an element to a superposition of pure tones is not the sum of the reactions of the element to single tones. In other words: Fourier-analysis fails here, the system being non-linear.

Therefore, in order to study the nervous activity evoked by vowel-like sounds we must stimulate the ear by means of a repeated damped oscillation the time-course of which is exactly known. We are not allowed, at this stage, to call this sound a vowel, because that would already be an interpretation.

It is evident that many investigators try to avoid the procedure just-sketched because it requires an advanced operation technique, the making and positioning of micro-electrodes, the means of making the given sound wave and the impulses it evokes in the nerve fibres simultaneously visible, and perhaps other intervening phenomena.

Some investigators concentrate on experiments with talking machines. Though talking machines may considerably differ in principle and construction they are all alike in that they possess *controls* that can be adjusted automatically by a programme or by hand (fig. 22).

The programme may be presented to the machine on punched cards, painted slides, rotating belts etc. By trial and error, sometimes guided by the results of some form of analysis, one arrives at a "best" setting of the controls. The best setting is the setting for which a group of listeners displays the highest degree of agreement in classifying a synthesized sound as a given phoneme.

Very much, however, depends on the instruction and previous

training given to the jury and on the number of phonemes from which the listeners are allowed to choose. When, for instance, the choice is one out of three even a deaf listener has a 33⅓% chance of being correct. The maker of the talking machine should never be in a jury because, like all parents, he understands his own child better than anyone else.

Now, properly speaking, with the ear as a *zero* instrument, we learn the properties of the talking machine rather than the properties of the ear. Though we are able to synthesize, in some way or another, a sound that a jury is willing to classify as one of a given number of phonemes, we have not thereby proved that the ear has a direct access to the properties the controls master, nor do we know whether the ear in actual speech uses other criteria not available on the machine. It is therefore, dangerous, to call the controls "the parameters of speech".

To study the mode of action of the ear without performing operations we need a different set-up. We need an artificial ear, in the ideal case consisting of a sound input and many nerve outputs (fig. 23). By listening to a given sound wave at the same time as seeing the impulses of the artificial ear on the screen of a cathode-ray oscilloscope, we use the brain as an instrument for comparing the output of our acoustic nerve to the impulses of the artificial nerve fibres. When we hear a change in the given sound wave we can see on the screen what change in the pattern of impulses was at the root of that change.

The construction of the artificial ear should be based on what we really know about the physiology of the organ of Corti and not on wishful thinking. We should not flirt with our supposed ignorance on that subject and we must not take refuge in mathematical ways of describing speech sounds like, for instance, Fourier-analysis, just to do something. We must not misuse the spectrograph.

Let us first enumerate items in favour of the spectrograph.

A good spectrograph is a valuable technical achievement. Though it does not depict phase it is still very convenient for those who study the vocal tract by investigating how the latter reacts to sinusoidal waves, often called pure tones. The spectrograph is very

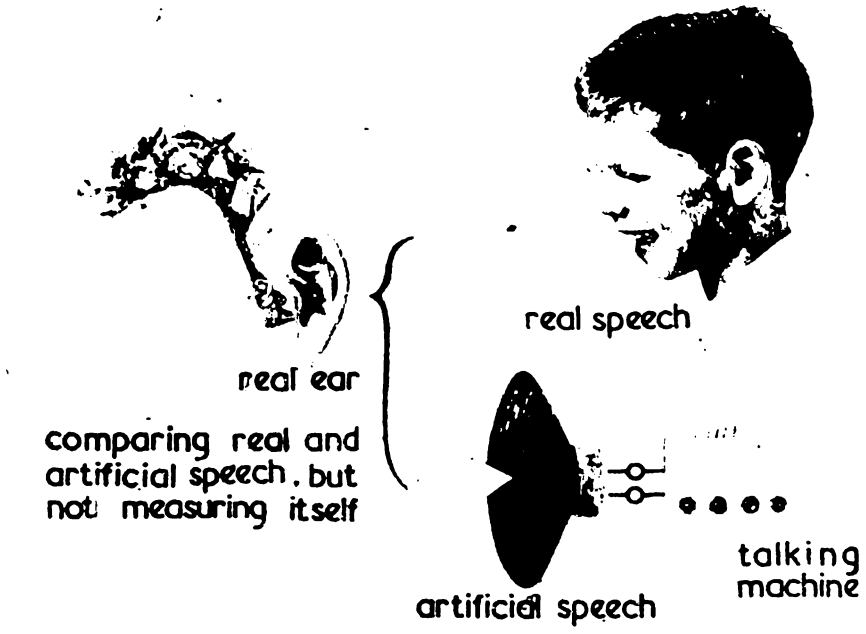


Fig. 22. Schematic representation of the process of setting the controls of a talking machine.

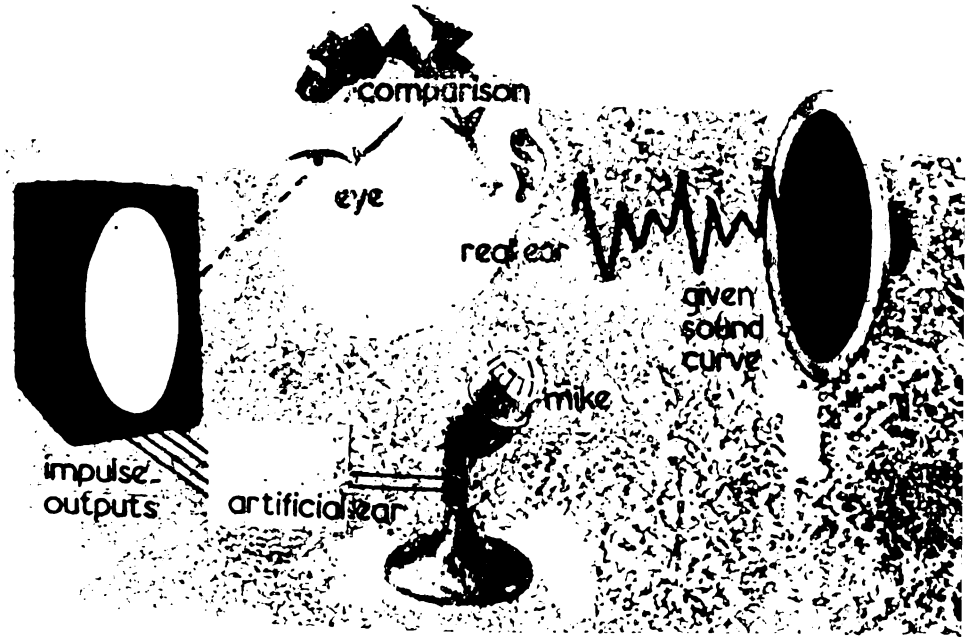


Fig. 23. A method for studying the mode of action of the ear by using a second sense organ. It mu
be emphasized here that the action of the brain is *highly* schematized in this figure, but that do
not reduce the value of the method.

helpful for comparing the output of an artificial vocal tract to that of the real mouth, especially when the artificial vocal tract is based on spectrographic principles.

We must not, however, apply the spectrograph in the study of the organ of hearing.

In this connection it is interesting to read what Dr. G. Fant states on page 160 of his excellent book *Acoustic Theory of Speech Production:* "It has become the normal technique in speech analysis to illustrate acoustic sound quality attributes by spectral curves. However, oscillograms produced at a high paper speed may sometimes provide a comparable or even clearer insight into specific details of the signal structure."

After a critical study of what has been written about aural stimuli and their interpretation I came to the conclusion that progress in this field is retarded by superstitions.

For instance, more often than not one reads in papers and textbooks that frequency, intensity and time *are* the physical components or parameters of speech sounds, as if the Fourier-components were physical realities.

The majority of people saying loose things about the ear assert that it is looking for those alleged physical components.

Now the only physical reality of a speech sound is the time function depicting the barometric pressure as a function of time. In its graphical form it is often called sound curve or oscillogram.

Physical realities are, for instance, the zero-crossings and peaks in a sound curve because they are events that really happen in the vibrating air at the entrance of the ear. The nerve endings, however, have no direct access to an air-borne sound curve which has to penetrate the ear until it reaches the organ of Corti where it is transformed into an electrical phenomenon that stimulates the nerve endings. No doubt during these adventures the shape of the curve changes but in a way that can nowadays be both predicted and measured. The nerve endings react to the zero-crossings in the ultimate curve by which they are stimulated. The performance of groups of nerve endings must also be seen in this light. In other words, the whole pattern of nervous activity in the acoustic nerve

is governed by the ability of the nerve endings to indicate zero crossings.

When, in future, the art of predicting or actually measuring the nervous activity evoked by speech sounds has developed sufficiently, investigators will be in a better position. They will then be able to study the purely interpretative faculties of the listener apart from the actual nervous activity at his disposal.

It will then also be possible to settle the following question. As far as we can see at the moment the overlap between consonants is less pronounced than between the vowels because the articulatory freedom in vowels is greater. We have a hunch that in running speech the consonants form the cues for the identification of the words rather than the vowels which are, as it were, filled in by the listener on the basis of interpretation.

Summarizing, we can underline the following points.

The aural stimuli are those physical events in the air that can be transformed into patterns of nervous activity by the ear.

The nature of this transformation must be studied further by means of experiments on real ears involving the measurement of the nervous activity evoked by known sound waves.

In current speech the nervous activity does not yield a one-to-one description of the vowel phonemes intended by the speaker.

A human listener is superior to a machine in that he can interpret by using his brain. Determining the degree in which he has to use this faculty in practice is a problem that can only be solved by close cooperation between linguists, psychologists, physiologists, and phoneticians.

FUNDAMENTALS OF FOURIER-ANALYSIS

When there is a time function $f(t)$, graphically represented by the curve in fig. 24, it is possible to write that function as follows:

$$f(t) = c_o + \sum_{n=1}^{n=\infty} a_n \sin n \frac{2\pi}{T} t + \sum_{n=1}^{n=\infty} b_n \cos n \frac{2\pi}{T} t \qquad (1)$$

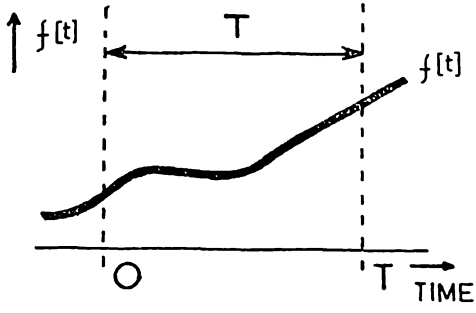


Fig. 24. A time function $f(t)$ in which we are interested only during the interval T.

where the coefficients c_o, a_n and b_n are given by:

$$c_o = \frac{1}{T} \int_o^T f(t)\, dt \qquad (2)$$

$$a_n = \frac{1}{2T} \int_o^T f(t) \sin n \frac{2\pi}{T} t\, dt \qquad (3)$$

$$b_n = \frac{1}{2T} \int_o^T f(t) \cos n \frac{2\pi}{T} t\, dt \qquad (4)$$

It is possible, to write equation (1) in a somewhat more simple form as follows:

$$f(t) = c_o + \sum_{n=1}^{n=\infty} c_n \sin\left(n \frac{2\pi}{T} t + \varphi_n\right) \tag{5}$$

where

$$c_n = \sqrt{a_n^2 + b_n^2} \tag{6}$$

and

$$tg\ \varphi_n = \frac{b_n}{a_n} \tag{7}$$

The proofs and limitations of these theorems can be found in any good textbook on Fourier-analysis.

Equation (5) embodies the traditional formulation:

Within the interval T the time function $f(t)$ may be regarded mathematically as the sum (or superposition) of an infinite number of sinusoidal waves the frequencies of which are *multiples* of the value $f_o = \frac{1}{T}$ which is called the fundamental frequency.

For that reason this type of analysis is called harmonic analysis. The thus-*defined* components are called harmonics. Each harmonic has its own frequency, its own amplitude (represented by c_n) and its own so-called phase-angle (symbolized by φ_n).

Before discussing in more detail the merits of Fourier-analysis itself we draw attention to the fact that it represents only one of the many possibilities of decomposing curves into a series of other curves of known mathematical description. It belongs to the class of developments into orthogonal functions. A decomposition into, for instance, Bessel functions, or even into square waves is also possible. There is nothing "physical" or "natural" about Fourier-analysis. It is merely an attractive possibility.

The Fourier series given by equation (5) only describes the function $f(t)$ *in the interval* T. What we do get, however, when we, as it were, forget that restriction, can be seen in fig. 25.

Owing to the periodic character of the sinusoidal components the Fourier series repeats itself and consequently the part of $f(t)$ bounded by the times o and T periodically at time intervals T.

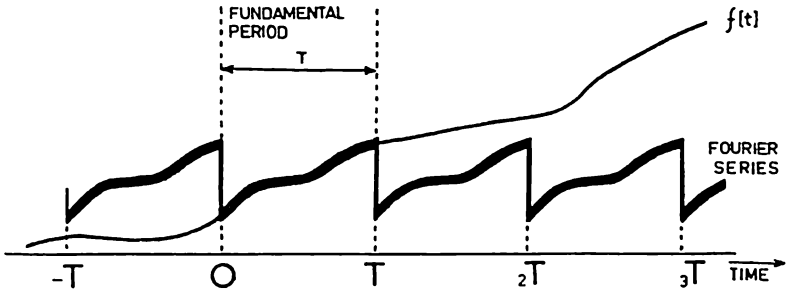


Fig. 25. Showing how the Fourier series with the fundamental period T periodically repeats $f(t)$ at intervals of T. This repetition is a mathematical extra but has nothing to do with the real time course of $f(t)$ outside the fundamental period during which one is interested in $f(t)$.

Outside the fundamental period of interest the Fourier-series does not describe $f(t)$ as will be clear from the figure.

The periodic character of the Fourier-series can be used to advantage in elegantly describing time functions that are really periodic in nature, provided we let the fundamental period of the Fourier-series coincide with the fundamental period of the $f(t)$ to be described. This restriction seems to be self-evident but fig. 26 shows that it is not.

There is nothing but common sense to forbid one to choose T arbitrarily. There is no obligation, however, to let T coincide with T_o.

With the arbitrary character of T in mind one must ask oneself how it is possible to construct a spectrograph (21, 22, 23, 24, 25, 26, 27) said to perform a frequency analysis. The answer is, that only in certain circumstances does the spectrograph produce part of a Fourier analysis, in spite of its often ingenious construction.

In the first place it can only deal adequately with periodic time functions. Suppose such a function has the fundamental period T_o.

Depending on its type a spectrograph contains one or more filters through which the periodic time function is passed. The filters are called band filters because they are designed to transmit only a limited range or band of frequencies, called the band width. When the band width is smaller than $f_o = \frac{1}{T_o}$ the filter will react

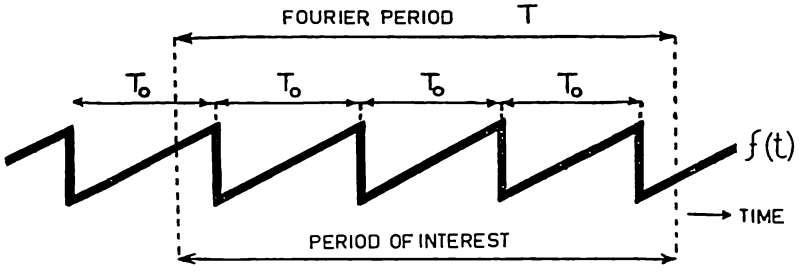


Fig. 26. The choice of the fundamental period T of the Fourier series. It need not necessarily coincide with the fundamental period T_0 of the periodic $f(t)$ to be described.

to the periodic function, as can be shown through calculation, by allowing only one sine wave to pass, the frequency of which is a multiple of f_0 that fits into the bandwidth. In theory the filter produces a pure sine wave only after the periodic function has been presented to it for an infinitely long time. For all practical purposes the filter will reach its steady state earlier, but the smaller the band width chosen the more time the filter needs to produce its result.

A filter with a band width of, for instance, 50 Hz (c/s) has a time constant of $\frac{1}{50}$ sec=20 m sec which is too high for many purposes.

In practice the filters of so-called broad band spectrographs have band widths of the order of 200 Hz (c/s) or even higher. Consequently they let pass not only one but even two Fourier-components for male voices. On the average the male voice produces vowels, that is to say periodic time-functions with a period of $\frac{1}{150}$ sec, corresponding to a repetition rate of 150 glottal impulses per second.

Normally, a spectrograph does not indicate the amplitudes of all Fourier-components.[1] It merely emphasizes the output of a filter showing more activity than its neighbours. This mode of action is inspired on the ideas of von Helmholtz who defined the formants as harmonic components of the glottal sound source that were amplified by the cavities of the vocal tract. This definition of the notion "formant" is fundamentally different from the original

[1] There are spectrographs, however, that can really portray all amplitudes in so-called sections.

definition of Hermann, who defined the formants as the frequencies of the damped oscillations set up in the vocal tract by the air puffs the vocal cords project into the pharynx.

As the amplitudes of the higher formants are much smaller than that of the first formant the higher harmonics are given a pre-amplification before feeding the sound into the filters. Also automatic gain control is applied (24) in order to compensate for the limited dynamic range of the recording paper. These measures render the spectrograph very susceptible to noise from tape recordings.

As the filters normally have a relatively large band width their output need not contain only one harmonic. Often two or more adjacent harmonics are passed, so that the output shows beats visible as vertical striations in the spectrogram (24). The striations betray the fundamental frequency of the voice and its changes without actually offering an adequate method of pitch recording. The spectrograph has not been designed for that purpose. The spectrograph does not and cannot show the phase-angle of the harmonics. As long as one confines its use to vowel-studies this does not seem to be a serious drawback.

When, for instance, explosive consonants are presented to the filters, the spectrograph will no doubt produce a reaction visible in the spectrogram.

These responses, however, cannot be related to Fourier analysis in the same simple way as in the case of sustained vowels.

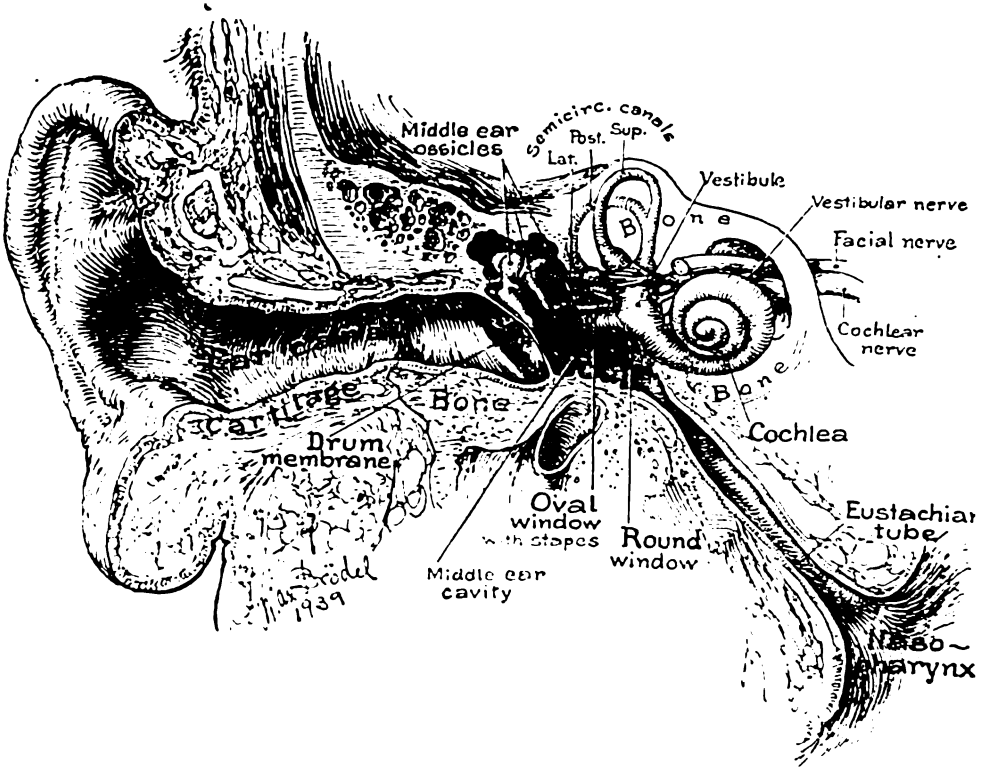


Fig. 27. Cross-section through the human ear. Drawing by Max Brödel. From L. G. Wever and M. Lawrence, *Physiological Acoustics* (Princeton, N.J., n.d.).

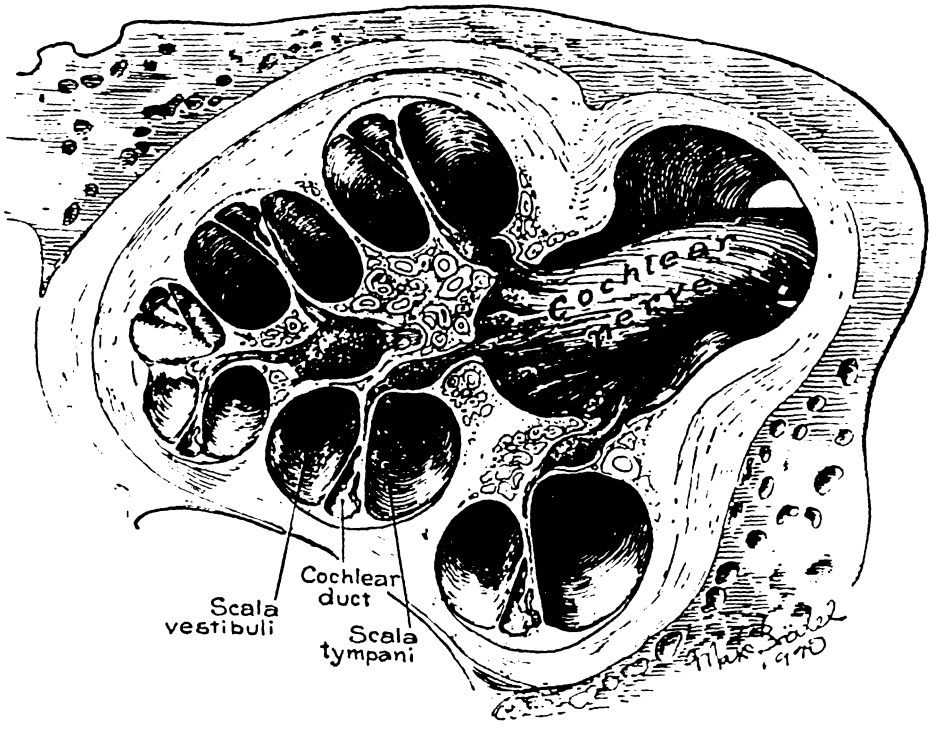
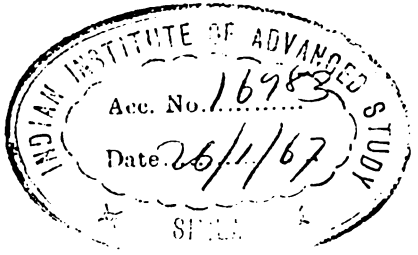


Fig. 28. A midmodiolar section of the cochlea. Drawing by Max Brödel. From the 1940 *Year Book of the Eye, Ear, Nose and Throat.*

REFERENCES

1. I. Tasaki, "Nerve impulses in individual auditory nerve fibres of Guinea pig," *J. Neurophysiol.*, 17 (1954), p. 97–122.
2. I. Tasaki and Hallowell Davis, "Electric responses of individual nerve elements in cochlear nucleus to sound stimulation (Guinea pig)," *J. Neurophysiol.*, 18 (1955), p. 151–158.
3. Robert Galambos, "Neural Mechanisms of Audition," *Physiological Reviews*, Vol. 34, No. 3 (July, 1954).
4. G. von Békésy, "Resonance curve and the decay period at various points on the cochlear partition," *JASA*, 21 (1949), p. 245.
5. *Lingua*, IV, 2 (1954), p. 167–193.
6. H. Mol and E. M. Uhlenbeck, "Hearing and the concept of the phoneme," *Lingua*, VIII, 2 (1959), p. 161–185.
7. H. Mol, *Belief and superstition in phonetics.* Inaugural Address, Amsterdam, 1960. [In Dutch.]
8. H. Fletcher, *Speech and hearing in communication*, p. 111.
9. H. Fletcher, *The mechanism of hearing as revealed through experiment on the masking effect of thermal noise* (= *Bell Telephone System, Technical Publications, Monogr. B-1091*).
10. G. von Békésy, "Microphonics produced by touching the cochlear partition with a vibrating electrode," *JASA*, 23 (1951), p. 29.
11. Engström, H. and J. Wersäll, *Acta Oto-Laryng.* 43 (1953), p. 323–334.
12. C. Wansdronk, *On the mechanism of hearing*, Thesis Leiden, November 1961.
13. H. Fletcher, *Speech and Hearing in Communication* (New York, 1953), ch. 14.
14. I. Zwislocki, *JASA*, 22 (1950), p. 778–784.
15. L. C. Peterson and B. F. Bogert, *JASA*, 22 (1950), p. 369–381.
16. Denker, A., *Das Gehörorgan und die Sprechwerkzeuge der Papageien* (Wiesbaden, 1907).
17. Landois, L., *Lehrbuch der Physiologie des Menschen* (Wien-Leipzig, 1889).
18. G. E. Peterson and H. L. Barney, *JASA*, 24 (1952), p. 175.
19. P. Ladefoged and D. E. Broadbent, "Information conveyed by vowels," *JASA*, 29 (1957), p. 98–104.
20. D. E. Broadbent, "The perception of speech," *Nature*, 189 (Febr. 18, 1961), no. 4764, p. 528–529.
21. Ernst Pulgram, *Introduction to the Spectrography of Speech* ('s-Gravenhage, 1959).
22. Potter, R. K., "Introduction to Technical Discussions of Sound Portrayal," *JASA*, 17 (1946), p. 1–3.

23. Steinberg, J. C. and N. R. French, "The Portrayal of Visible Speech," *JASA*, 17 (1946), p. 4–18.
24. Koenig, W., H. K. Dunn, and L. Y. Lacy, "The Sound Spectrograph," *JASA*, 17 (1946), p. 19–49.
25. Riesz, R. R., and L. Schott, "Visible Speech Cathode-Ray Translator," *JASA*, 17 (1946), p. 50–61.
26. Dudley, H. W., and O. O. Gruenz, Jr., "Visible Speech Translators with External Phosphors," *JASA*, 17 (1946), p. 62–73.
27. Kopp, G. A., and Harriet C. Green, "Basic Phonetic Principles of Visible Speech," *JASA*, 17 (1946), p. 74–89.



JANUA LINGUARUM

SERIES MINOR

1. ROMAN JAKOBSON and MORRIS HALLE: Fundamentals of Language. 1956. 97 pp. Gld. 6.—
3. EMIL PETROVICI: Kann das Phonemsystem einer Sprache durch fremden Einfluss umgestaltet werden? Zum slavischen Einfluss auf das rumänische Lautsystem. 1957. 44 pp. Gld. 4.—
4. NOAM CHOMSKY: Syntactic Structures. Third printing, 1963. 118 pp. Gld. 8.—
5. N. VAN WIJK: Die baltischen und slavischen Akzent- und Intonationssysteme. Ein Beitrag zur Erforschung der baltisch-slavischen Verwandtschaftsverhältnisse. 2nd ed. 1958. 160 pp. Gld. 15.—
6. BERNARD GEIGER, TIBOR HALASI-KUN, AERT H. KUIPERS and KARL H. MENGES: Peoples and Languages of the Caucasus. A Synopsis. 1959. 77 pp., map. Gld. 8.—
7. ERNST PULGRAM: Introduction to the Spectrography of Speech. 1959. 174 pp., 31 figs., 2 tables. Gld. 12.—
8. AERT H. KUIPERS: Phoneme and Morpheme in Kabardian (Eastern Adyghe). 1960. 124 pp. Gld. 16.—
10. URIEL and BEATRICE WEINREICH: Yiddish Language and Folklore. A Selective Bibliography for Research. 1959. 66 pp. Gld. 6.—
11. E. and K. DELAVENAY: Bibliography of Mechanical Translation. – Bibliographie de la traduction automatique. 1960. 69 pp. Gld. 10.—
12. CARL L. EBELING: Linguistic Units. Second printing 1962. 143 pp. Gld. 12.—
13. SHELOMO MORAG: The Vocalization Systems of Arabic, Hebrew and Aramaic. Their Phonetic and Phonemic Principles. 1962. 85 pp., two folding tables. Gld. 15.—
14. DWIGHT L. BOLINGER: Generality, Gradience, and the All-or-None. 1961. 46 pp. Gld. 5.50
15. ALPHONSE JUILLAND: Outline of a General Theory of Structural Relations. 1961. 58 pp. Gld. 7.50
16. Sens et usages du terme Structure, dans les sciences humaines et sociales. Édité par ROGER BASTIDE. 1962. 165 pp. Gld. 16.—
17. W. SIDNEY ALLEN: Sandhi. The Theoretical, Phonetic, and Historic Bases of Word-Junction in Sanskrit. 1962. 114 pp. Gld. 16.—

19. WALERIAN ŚWIECZKOWSKI: Word Order Patterning in Middle English. A quantitative study based on *Piers Plowman* and *Middle English Sermons*. 1962. 114 pp. Gld. 16.—

20. FINNGEIR HIORTH: Zur formalen Charakterisierung des Satzes. 1962. 152 pp. Gld. 15.—

22. E. F. HADEN, M. S. HAN and Y. W. HAN: A Resonance-Theory for Linguistics. 1962. 51 pp. Gld. 7.—

23. S. LEVIN: Linguistic Structures in Poetry. 1962. 64 pp. Gld. 8.—

24. ALPHONSE JUILLAND, and JAMES MACRIS: The English Verb System. 1962. 81 pp. Gld. 8.—

27. LÁSZLÓ ANTAL: Questions of Meaning. 1963. 95 pp. Gld. 10.—

SERIES MAIOR

2. DEAN S. WORTH: Kamchadal Texts collected by W. Jochelson. 1961. 284 pp. Cloth. Gld. 58.—

3. PETER HARTMANN: Theorie der Grammatik.

   2. Zur Konstitution einer allgemeinen Grammatik. 1961. 136 pp. Gld. 21.—

   3. Allgemeinste Strukturgesetze in Sprache und Grammatik. 1961. 136 pp. Gld. 21.—

   4. Grammatik und Grammatizität. 1962. 144 pp. Gld. 24.—

4. GUSTAV HERDAN: Type-Token Mathematics. A Textbook of Mathematical Linguistics. 1960. 448 pp., 17 figs. Cloth. Gld. 54.—

6. TATIANA SLAMA-CAZACU: Langage et Contexte. Le problème du langage dans la conception de l'expression et de l'interprétation par des organisations contextuelles. 1961. 251 pp., 5 figs. Cloth. Gld. 48.—

7. ALF SOMMERFELT: Diachronic and Synchronic Aspects of Language. Selected Articles. 1962. 421 pp., 23 figs. Cloth. Gld. 54.—

8. THOMAS A. SEBEOK and VALDIS J. ZEPS: Concordance and Thesaurus of Cheremis Poetic Language. 1961. 259 pp. Cloth. Gld. 58.—

9. GUSTAV HERDAN: The Calculus of Linguistic Observations. 1962. 271 pp., 6 figs. Cloth. Gld. 42.—

10. Proceedings of the Fourth International Congress of Phonetic Sciences, held at the University of Helsinki, 4–9 September 1961. Edited by ANTTI SOVIJÄRVI and PENTTI AALTO. 1962. 855 pp., num. figs. and plates. Cloth. Gld. 125.—

14. RUTH HIRSCH WEIR: Language in the Crib. 1962. 216 pp, Cloth. Gld. 32.—