good deal about efficient communication. They suggest possibilities for coding speech information in forms other than merely the transduced pressure wave. The normal mode and excitation relations, for example, indicate a schema on which an analysis-synthesis transmission system might be based. The same can be said for describing the vocal tract by articulatory parameters. Both results reflect constraints peculiar to the speech-producing mechanism.

As yet, however, the properties of hearing and the constraints exhibited by the ear have not entered the discussion. The next chapter proposes to establish certain fundamental properties of the mechanism of hearing – so far as they are known. The exposition will follow a pattern similar to that of the present chapter. The results of both fundamental discussions will then be useful in subsequent consideration of speech analysis and speech synthesis.

IV. The Ear and Hearing

The ultimate recipient of information in a speech communication link usually is man. His perceptual abilities dictate the precision with which speech data must be processed and transmitted. These abilities essentially prescribe fidelity criteria for reception and, in effect, determine the channel capacity necessary for the transmission of voice messages. It consequently is pertinent to inquire into the fundamental mechanism of hearing and to attempt to establish capabilities and limitations of human perception.

As suggested earlier, speech information – originating from a speaker, traversing a transmission medium and arriving at a listener – might be considered at a number of stages of coding. On the transmitter side, the stages might include the acoustic wave, the muscular forces manipulating the vocal mechanism, or the physical shape and excitation of the tract. On the receiver side, the information might be considered in terms of the acoustic-mechanical motions of the hearing transducer, or in terms of the electrical pulses transmitted to the brain over the auditory nerve. Characteristics of one or more of these codings might have application in practicable transmission systems.

The previous chapter set forth fundamental relations between the acoustics and the physiology of the vocal mechanism. We will subsequently have occasion to apply the results to analysis-synthesis telephony. In the present chapter we wish to establish similar relations for the ear. Later we will utilize these in discussions of auditory discrimination and speech perception.

4.1. Mechanism of the Ear

The acousto-mechanical operation of the peripheral ear has been put on a rather firm base. This knowledge is due primarily to the brilliant experiments carried out by G. VON BÉKÉSY, and for which he was awarded the Nobel Prize in 1961. In contrast, present knowledge is relatively incomplete about inner-ear processes for converting mechanical motion into neural activity. Still less is known about the transmission of neural information to the brain and the ultimate mechanism of perception.

Despite these difficulties, it is possible to quantify certain aspects of perception without knowing in detail what is going on inside the "black box". Subjective behavior, in response to prescribed auditory stimuli, can of course be observed and measured, and such data are useful guideposts in the design of speech communication systems. In some instances the correlations between perceptual behavior and the physiological operation of the peripheral ear can be placed in clear evidence. The present discussion aims to indicate current understanding of auditory physiology and psychoacoustic behavior, and to illustrate the extent to which the two can be brought into harmony.

The primary acoustic transducer of the human is shown schematically in Fig. 4.1. The acousto-mechanical components of the organ are conventionally divided according to three regions, namely, the outer ear, the middle ear, and the inner ear.

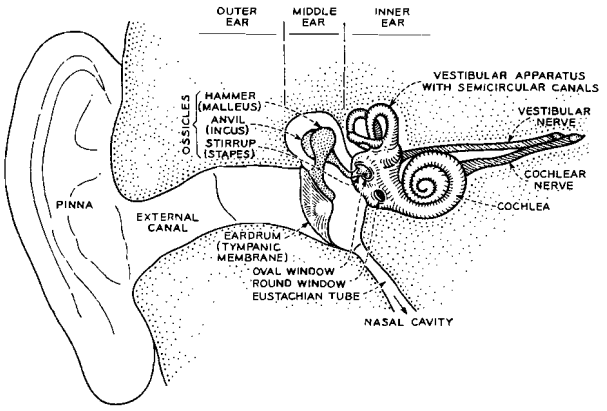

Fig. 4.1. Schematic diagram of the human ear showing outer, middle and inner regions. The drawing is not to scale. For illustrative purposes the inner and middle ear structures are shown enlarged

good deal about efficient communication. They suggest possibilities for coding speech information in forms other than merely the transduced pressure wave. The normal mode and excitation relations, for example, indicate a schema on which an analysis-synthesis transmission system might be based. The same can be said for describing the vocal tract by articulatory parameters. Both results reflect constraints peculiar to the speech-producing mechanism.

As yet, however, the properties of hearing and the constraints exhibited by the ear have not entered the discussion. The next chapter proposes to establish certain fundamental properties of the mechanism of hearing – so far as they are known. The exposition will follow a pattern similar to that of the present chapter. The results of both fundamental discussions will then be useful in subsequent consideration of speech analysis and speech synthesis.

IV. The Ear and Hearing

The ultimate recipient of information in a speech communication link usually is man. His perceptual abilities dictate the precision with which speech data must be processed and transmitted. These abilities essentially prescribe fidelity criteria for reception and, in effect, determine the channel capacity necessary for the transmission of voice messages. It consequently is pertinent to inquire into the fundamental mechanism of hearing and to attempt to establish capabilities and limitations of human perception.

As suggested earlier, speech information – originating from a speaker, traversing a transmission medium and arriving at a listener – might be considered at a number of stages of coding. On the transmitter side, the stages might include the acoustic wave, the muscular forces manipulating the vocal mechanism, or the physical shape and excitation of the tract. On the receiver side, the information might be considered in terms of the acoustic-mechanical motions of the hearing transducer, or in terms of the electrical pulses transmitted to the brain over the auditory nerve. Characteristics of one or more of these codings might have application in practicable transmission systems.

The previous chapter set forth fundamental relations between the acoustics and the physiology of the vocal mechanism. We will subsequently have occasion to apply the results to analysis-synthesis telephony. In the present chapter we wish to establish similar relations for the ear. Later we will utilize these in discussions of auditory discrimination and speech perception.

4.1. Mechanism of the Ear

The acousto-mechanical operation of the peripheral ear has been put on a rather firm base. This knowledge is due primarily to the brilliant experiments carried out by G. VON BÉKÉSY, and for which he was awarded the Nobel Prize in 1961. In contrast, present knowledge is relatively incomplete about inner-ear processes for converting mechanical motion into neural activity. Still less is known about the transmission of neural information to the brain and the ultimate mechanism of perception.

Despite these difficulties, it is possible to quantify certain aspects of perception without knowing in detail what is going on inside the "black box". Subjective behavior, in response to prescribed auditory stimuli, can of course be observed and measured, and such data are useful guideposts in the design of speech communication systems. In some instances the correlations between perceptual behavior and the physiological operation of the peripheral ear can be placed in clear evidence. The present discussion aims to indicate current understanding of auditory physiology and psychoacoustic behavior, and to illustrate the extent to which the two can be brought into harmony.

The primary acoustic transducer of the human is shown schematically in Fig. 4.1. The acousto-mechanical components of the organ are conventionally divided according to three regions, namely, the outer ear, the middle ear, and the inner ear.

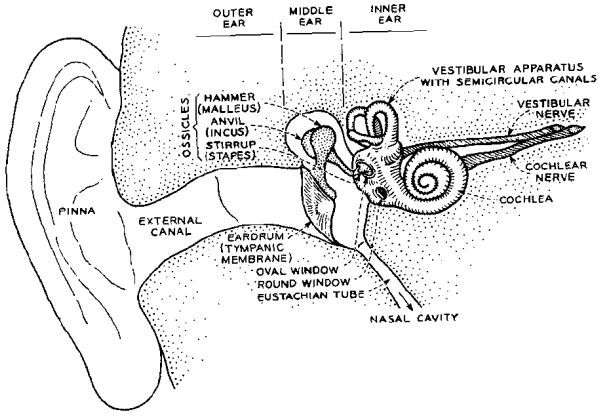

Fig. 4.1. Schematic diagram of the human ear showing outer, middle and inner regions. The drawing is not to scale. For illustrative purposes the inner and middle ear structures are shown enlarged

4.11. The Outer Ear

As commonly understood, the term *ear* usually applies to the salient, convoluted appendage on the side of the head. This structure is the pinna, and it surrounds the entrance to the external ear canal. Its main function in man is to protect the external canal—although its directional characteristics at high audible frequencies probably facilitate localization of sound sources. (In some animals, the directional acoustic properties of the pinna are utilized more fully.)

In man, the external ear canal, or meatus, is about 2.7 cm in length and about 0.7 cm in diameter. Its volume is on the order of 1 cm^3, and its cross-section is oval-to-circular in shape with an area 0.3 to 0.5 cm^2 (BÉKÉSY and ROSENBLITH; DAVIS, 1951). The meatus is terminated by a thin membrane which is the eardrum, or tympanic membrane. The membrane has the form of a relatively stiff, inwardly-directed cone with an included angle of about 135°. Its surface area is on the order of 0.8 cm^2. To a rough approximation, the meatus is a uniform pipe—open at one end and closed at the other. It has normal modes of vibration which occur at frequencies where the pipe length is an odd multiple of a quarter wavelength. The first mode therefore falls at $f \cong c/4(2.7) \cong$ 3000 cps. This resonance might be expected to aid the ear's sensitivity in this frequency range. Measurements do in fact show that it provides a sound pressure increase at the ear drum of between 5 and 10 db over the value at the canal entrance (WIENER and ROSS).

4.12. The Middle Ear

Just interior to the eardrum is the air-filled, middle-ear cavity which contains the ossicular bones. The function of the ossicles is mainly one of impedance transformation from the air medium of the outer ear to the liquid medium of the inner ear[1]. The malleus, or hammer, is fixed to and rests on the eardrum. It makes contact with the incus, or anvil, which in turn connects via a small joint to the stapes, or stirrup. The footplate of the stirrup seats in a port, the oval window, and is retained there by an annular ligament. The oval window is the entrance to the inner ear.

A sound wave impinging on the outer ear is led down the external meatus and sets the eardrum into vibration. The vibration is transmitted via the three ossicular bones into the inner ear. The acousto-mechanical impedance of the inner ear is much greater than that of air, and for efficient transmission of sound energy an impedance transformation (a step up) is required. The ossicles provide such. First their lever action

[1] This impedance transformation is important to the basic role of the middle ear; that is, the conversion of an external sound pressure into a fluid volume displacement in the inner ear (see Sec. 4.13).

alone provides a force amplification of about 1.3 (BÉKÉSY, 1960). That is, a force applied to the hammer appears at the stirrup footplate multiplied by 1.3. Second, the effective area of the eardrum is much greater than that of the stirrup, so that the ratio of pressure applied at the stirrup to that applied at the eardrum is essentially 1.3 times the ratio of the effective areas of drum and stirrup. BÉKÉSY has measured this pressure transformation and finds it to be on the order of 15:1.

The middle ear structure serves another important purpose, namely, it provides protection against loud sounds which may damage the more delicate inner ear. The protective function is generally assumed to be served by two tympanic muscles—especially the tensor-tympani which connects the middle of the eardrum to the inner region of the head. Reflex contractions presumably attenuate the vibratory amplitude of the drum. BÉKÉSY points out, however, that voluntary contractions of the tensor and changes in the static pressure of the meatus only slightly reduce the vibrational amplitude of the drum. The contractions consequently can have only small effect in protecting against sound pressures that extend over a wide range of magnitudes. This fact can be established from measurements of the acoustic impedance at the drum.

In detailed studies on the mode of vibration of the ossicles, BÉKÉSY observed that at low and moderate sound intensities the stapes motion is principally a rotation about an axis through the open "hoop" of the stirrup. The movement is illustrated in Fig. 4.2a. At sound intensities near and above the threshold of feeling, the motion of the stapes changes more to a rotation about an axis running longitudinally through the "arch" of the stapes, as shown in Fig. 4.2b. In the latter mode, the effective volume displacement is small because the upper-half of the footplate advances by about as much as the lower half recedes.

Contraction of the middle ear muscles increases with sound intensity, so that the ossicles are prevented from bouncing out of contact and causing excessive distortion at the high levels. This control of distortion

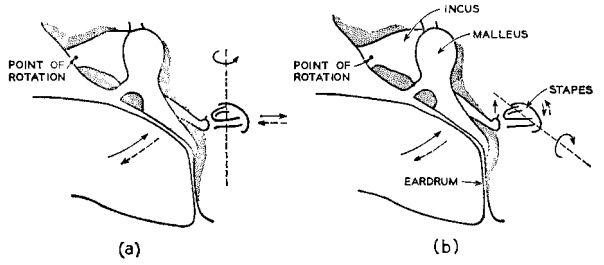

Fig. 4.2a and b. Vibration modes of the ossicles. (a) sound intensities below threshold of feeling (b) intensities above threshold of feeling. (After BÉKÉSY, 1960)

over the amplitude range from threshold-of-hearing to near threshold-of-feeling—while at the same time protecting the inner ear from harmful vibrational levels—apparently accounts for the elaborate middle-ear structure[1].

One of the important characteristics of the middle ear is its transmission as a function of frequency, that is, the volume displacement of the stapes footplate produced by a given sound pressure at the eardrum. A number of efforts have been made to measure or to deduce this characteristic (BÉKÉSY, 1960; Zwislocki, 1957, 1959; MØLLER, 1961, 1962). The results are somewhat disparate, suggesting that not only is the characteristic a function of intensity in the living human, but that it may vary substantially from individual to individual.

If the fluid of the inner ear is considered incompressible and the walls of the cochlea rigid, then the volume displacement of the round window must be the same as that of the stapes footplate. At low frequencies the combined elasticity of the drum, ossicles and round window membrane controls the stirrup motion. That is, the system acts like a spring, with the stapes displacement proportional to, and in phase with, the eardrum pressure. Somewhere between about 1000 and 3000 cps the mass reactance of the system becomes important, and the motion passes from a stiffness-controlled vibration to a viscous-controlled one and finally to a mass-controlled motion. For a given sound pressure at the drum, the stirrup displacement begins to diminish in amplitude and lag in phase as frequency increases.

BÉKÉSY (1960) has made a number of measurements of middle-ear transmission by directly observing the volume displacement of the round window. The transmission properties can also be deduced from a knowledge of the middle-ear topology, the input mechanical impedance to the inner ear, and the acoustic impedance at the eardrum. This approach has been used by ZWISLOCKI (1957, 1959) and by MØLLER (1961) to develop analog circuits of the middle ear. All these results agree in gross aspects but suggest that substantial variability can exist in the characteristic. By way of comparison, the transmission of the middle ear according to several determinations is shown in Fig. 4.3a–d.

For the data in Fig. 4.3b, BÉKÉSY obtains a critical "roll-off" frequency for middle-ear transmission of about 800 cps. For the data in Fig. 4.3a, it is clearly higher, possibly around 3000 cps. ZWISLOCKI's result in Fig. 4.3c places it somewhere in the vicinity of 1500 cps, and

[1] One can appreciate the difficulties posed in duplicating this mechanical linkage with prosthetic devices. For example, one middle-ear prosthesis involves replacing damaged or diseased ossicles by a plastic strut joining the drum and the stapes footplate. The protection against distortion and high-amplitude vibration, normally provided by the middle ear, are difficult to include in such a construction.

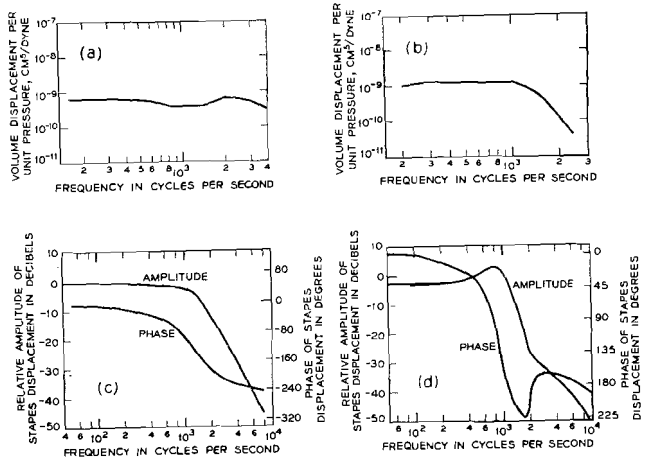

Fig. 4.3a–d. Data on middle ear transmission; effective stapes displacement for a constant sound pressure at the eardrum. (a) BÉKÉSY (1960) (one determination); (b) BÉKÉSY (1960) (another determination); (c) measured from an electrical analog circuit (after ZWISLOCKI, 1959); (d) measured from an electrical analog circuit (after MØLLER, 1961)

MØLLER's result in Fig. 4.3d is near 1000 cps. The common indication is that the middle-ear transmission has a low-pass characteristic. The effective cut-off frequency and the skirt slope are apparently subject to considerable variation.

4.13. The Inner Ear

As illustrated in Fig. 4.1, the inner ear is composed of the cochlea (normally coiled like a snail shell in a flat spiral of two and one-half turns), the vestibular apparatus and the auditory nerve terminations. It is in the cochlea that auditory mechanical-to-neural transduction takes place. The vestibular components (semi-circular canals, saccule and utricle) serve the sense of spatial orientation and apparently are not normally used for detecting audio vibrations.

If the cochlea is uncoiled and stretched out, it appears schematically as in Fig. 4.4. The cochlear chamber is filled with a colorless liquid, perilymph, which has a viscosity about twice that of water and a specific gravity of about 1.03. The length of the canal in the spiral conch is about 35 mm. The cross-sectional area at the stirrup end is about 4 mm² and the area decreases to about 1 mm² at the tip.

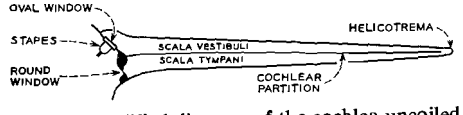

Fig. 4.4. Simplified diagram of the cochlea uncoiled

The cochlear chamber is divided along almost its whole length by a partition. The half which receives the stapes is called the scala vestibuli; the other half is the scala tympani. The cochlear partition is itself a channel – the scala media – bounded partly by a bony shelf, a gelatinous membrane called the basilar membrane, and another membrane known as REISSNER's membrane. The partition is filled with a different liquid, the endolymph. The basilar membrane and bony shelf both terminate a mm or two short of the ends of the scalas, permitting them to communicate at the helicotrema. The area of the connecting passage is about 0.3 to 0.4 mm^2 (BÉKÉSY and ROSENBLITH). The basilar membrane is about 32 mm in length and tapers from a width of about 0.05 mm at the base (stirrup) to about 0.5 mm at the apex (DAVIS, 1951).

The inner ear is connected to the middle ear at the stapes footplate. The latter, supported by a ring-shaped ligament, seats into the oval window (about 3 mm^2 in area). In vibrating, the stapes acts as a piston and produces a volume displacement of the cochlear fluid. Because the cochlea is essentially rigid and its fluid incompressible, fluid displacements caused by inward motion of the stapes must be relieved. This function is accomplished at the round window which is covered by a compliant membrane (about 2 mm^2). Very slow vibrations of the stapes (say less than 20 cps) result in a to-and-fro flow of fluid between the scala vestibuli and scala tympani through the opening at the helicotrema. Higher frequency vibrations are transmitted through the yielding cochlear partition at a point which depends uqon the frepuency content of the stimsound ulation.

A cross-section of the cochlea and its partition is shown in Fig. 4.5. The main functions and dynamical properties of the partition reside in

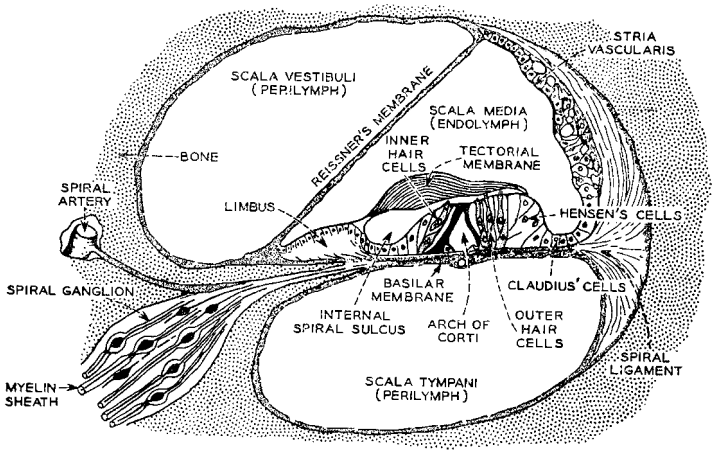

Fig. 4.5. Schematic cross section of the cochlear canal. (Adapted from DAVIS, 1957)

the basilar membrane. It is upon the latter that the organ of Corti rests. Among several types of supporting cells, the organ of Corti contains some 30000 sensory cells (or hair cells), on which the endings of the auditory nerve (entering from the lower left in Fig. 4.5) terminate. The basilar membrane is stiffer and less massive at its narrow, basal end and more compliant and massive at the broad, apical end. Its resonant properties therefore vary continuously along its length. At low frequencies, REISSNER'S membrane normally moves cophasically with the basilar membrane.

Current knowledge of the acoustic-mechanical properties of the basilar membrane is due almost exclusively to the efforts of VON BÉKÉSY. In physiological preparations, he vibrated the stapes footplate sinusoidally and measured the amplitude and phase of the membrane displacements along the length of the cochlea. The mechanical characteristics of the basilar membrane, as determined in these experiments, are shown in Fig. 4.6. Figs. 4.6a and b show the amplitude and phase of specific membrane points as functions of frequency. Fig. 4.6c shows the amplitude and phase as a function of membrane place with frequency as the parameter.

The amplitude and phase response of a given membrane point is much like that of a relatively broad band-pass filter. The amplitude responses of successive points are roughly constant-Q in nature. Because of this constant percentage bandwidth property, the frequency resolution is best at the low-frequency (apical) end of the membrane, and the time resolution is best at the higher-frequency (basal) end [1].

All the amplitude responses of Fig. 4.6 are normalized to unity. BÉKÉSY's measurements suggest, however, that for constant amplitude of stapes displacement, the peak membrane response increases at about 5 db/octave for points resonant at frequencies up to about 1000 cps, and is approximately constant in peak displacement for points resonant at higher frequencies. Linear increments of distance along the basilar membrane correspond approximately to logarithmic increments of peak frequency, at least for frequencies less than about 1000 cps.

Excitation at the stapes is propagated down the membrane in the form of a travelling wave of displacement. Because of the taper of the distributed constants with distance, essentially no reflection takes place at the helicotrema, and no standing wave of displacement is created. The membrane is a dispersive transmission medium. The travelling wave loses more and more of its high frequency components as it progresses toward the helicotrema, and its group delay increases.

[1] Recent measurements of basilar membrane vibration in animals, using the Mössbauer effect (JOHNSTONE and BOYLE; RHODE), suggest that the mechanical response is sharper (higher in Q) than shown in Fig. 4.6. Also, the measurements suggest that the mechanical response is somewhat dependent upon sound intensity.

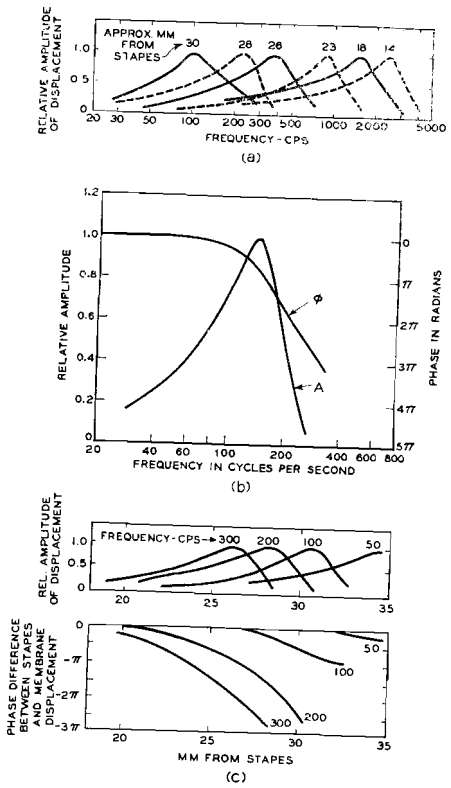

Fig. 4.6a–c. Amplitude and phase responses for basilar membrane displacement. The stapes is driven sinusoidally with constant amplitude of displacement. (After BÉKÉSY, 1960.) (a) Amplitude *vs* frequency responses for successive points along the membrane. (b) Amplitude and phase responses for the membrane place maximally responsive to 150 cps. (c) Amplitude and phase of membrane displacement as a function of distance along the membrane. Frequency is the parameter

4.14. Mechanical-to-Neural Transduction

Mechanical motion of the membrane is converted into neural activity in the organ of Corti. An enlarged view of this structure is shown in Fig. 4.7. The organ of Corti contains a collection of cells among which are the hair cells. The hairs emanating from these sensory cells protrude upward through the reticular lamina and contact a third membrane of the cochlear partition, the tectorial membrane. One set of cells lies in a single row, longitudinal to the basilar membrane and toward the axis of the cochlear spiral (left of the arch of Corti). They are termed the inner hair cells. Another set lies in three or four longitudinal rows, radially away from the center of the spiral. These are the outer hair

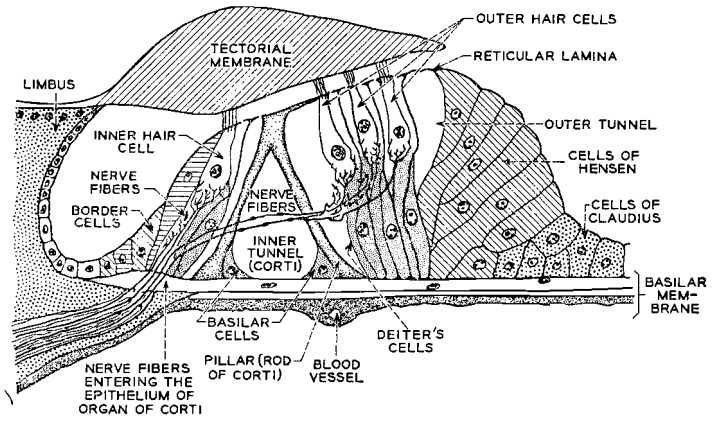

Fig. 4.7. Cross section of the organ of Corti. (After DAVIS, 1951)

cells. Estimates fix the number of the former at about 5000 and the latter at about 25000.

The tectorial and basilar membranes are anchored at their inner edges at spatially separate points. A deformation of the basilar membrane causes relative motion between the tectorial membrane and the reticular lamina and a resultant stress on the hairs passing between. By a process that presently is not understood, a bending of the hairs produces an electrical discharge in the cochlear portion of the VIIIth nerve[1]. The first-order fibers of this cochlear branch, or auditory nerve, enter from the lower left in Fig. 4.7 and run to the inner and outer hair cells.

Electrophysiological experiments suggest that the outer and inner hair cells of the organ of Corti differ in their sensitivities to mechanical stimulation (BÉKÉSY, 1953; DAVIS, 1958). The outer hair cells appear to be sensitive to bending only in a direction transverse to the long dimension of the membrane. Moreover, only outward bending of the hairs (away from the arch of Corti) produces an electrical potential in the scala media favorable for exciting the auditory nerve endings. This outward bending is produced on (unipolar) upward motions of the basilar membrane, that is, motions which drive it toward the tectorial membrane.

The inner hair cells, on the other hand—residing between the arch of Corti and the axis of the cochlear spiral—appear sensitive to bending in a direction parallel to the long dimension of the membrane (BÉKÉSY, 1953; DAVIS, 1958). In this case bending only toward the apex of the cochlea produces a scala media potential favorable for stimulating the

[1] The VIIIth nerve also serves the vestibular apparatus. See Fig. 4.1.

nerve. So far as a given point on the membrane is concerned, the inner hair cells are essentially sensitive to the longitudinal gradient of displacement, that is, to the spatial derivative in the long dimension. Furthermore, the inner cells fire only on the polarity of the gradient which corresponds to bending toward the apex. The threshold for firing of the inner cells appears to be appreciably higher than that for the outer cells. Exactly how the pattern of mechanical displacement of the basilar membrane is reflected in the "transducer" potentials of the sensory cells and in the electrical pulses elicited in the auditory nerve has yet to be put on a firm basis.

The sensory cells of the ear connect to the brain via the bundle of nerve cells – or neurons – comprising the auditory nerve. The auditory nerve passes down the axis of the cochlear spiral – collecting more nerve fibers as it runs from apex to base – until it contains some 30000 neurons. Neurons presumably have only two states, namely, active or inactive. When excited by an electrical input above a particular threshold, they produce a standard electrical pulse of about a millisecond duration and are desensitized for a period of about one to three milliseconds thereafter. They consequently can be excited to maximum discharge rates on the order of 300 to 1000 sec^{-1}.

The connections between the nerve cells and the hair cells in the organ of Corti are complex. Each inner hair cell is innervated by one or two nerve fibers, and each fiber connects with one or two hair cells. Innervation of the outer cells is more compound. Most nerve fibers make connections with a number of outer cells, and each outer cell usually receives connections from several nerve fibers (DAVIS, 1957). The exact functional significance of this complex multiple distribution of the nerve supply is not presently known. One study has suggested that it contributes to the great intensity range of the ear (VAN BERGEIJK).

The fibers of the auditory nerve twist like strands of rope about a central core. The nerve itself is short and enters the lower brain stem (medulla oblongata) after a run of about 5 mm (DAVIS, 1957). The incoming fibers divide, and the branches run respectively to the dorsal and to the vertral portions of the cochlear nucleus. Here the first synapses (junctions connecting one nerve cell to another) of the auditory system reside. The fibers of the auditory nerve, and the cells of the cochlear nucleus to which they join, essentially preserve the orderly arrangement of the corresponding sensory cells on the basilar membrane. The same general tendency toward orderly arrangement, with respect to membrane place of origin, seems to be maintained throughout the auditory system.

Relatively little is known about the mechanism by which the basilar membrane displacements are converted into neural activity. Still less is known about how information is coded in nerve pulses and assimilated into an auditory percept by the brain. Qualitatively, however, several deductions seem to be generally accepted. First, the hairs of the sensory cells, in experiencing a lateral shear owing to relative motion of basilar membrane and tectorial membrane (see Fig. 4.7), generate local electrical potentials which represent the local basilar membrane displacement. More precisely, the shearing forces on the sensory hairs "modulate" (as would a variable resistor) a current passing between the scala media and the base of the hair cell (DAVIS, 1965).

Second, this facsimile alternating potential, acting at the base of the hair cell, modulates the liberation of a chemical mediator about some quiescent rate. The mediator, in sufficient quantity, stimulates the dendritic endings of the first-order nerve fibers and causes the fibers to fire. Because of its quiescent bias rate, the hypothesized chemical mediator is secreted more on one phase of the sensory potential than on the other; that is, a rectifying function is implied in triggering the nerve fiber.

Lastly, the chemical stimulation of the nerve endings produces an all-or-none electrical firing, which is propagated axonally to subsequent higher-order fibers in the central nervous system.

There are two basic electrical phenomena in the cochlea: the resting (dc) polarization of the parts, and the ac output of the organ of Corti (which, as stated, appears to reflect the displacement of the cochlear partition). Current thinking holds that the ac output, or the cochlear microphonic[1], is a potential produced by the sensory or receptor cells and is derived from the pre-existing resting polarization of the receptor cells by a change in the ohmic resistance of a mechanically-sensitive portion of the cell membrane. "This change in resistance is presumably brought about by a deformation, however slight, of a critical bit of the polarized surface" (DAVIS, 1965).

Energy considerations argue for an active (power-amplifying) type of transduction, as in a carbon microphone. The biasing current (power supply) for the transducer is produced by the biological battery which is the resting polarization of the hair cell. The mechanical energy of the basilar membrane is not transduced into electrical current, rather it controls or modulates the flow of current across the interface (cell membrane) which separates the negative polarization inside the hair cell from the positive, endo-cochlear potential of the endolymph inside the cochlear partition.

A map of the cochlear resting potentials is shown in Fig. 4.8. The scala tympani is taken as the zero reference potential, and regions of similar potential are often found within the organ of Corti. Other areas,

[1] This potential is typically observed by an electrode placed at the round window or inserted into a scala.

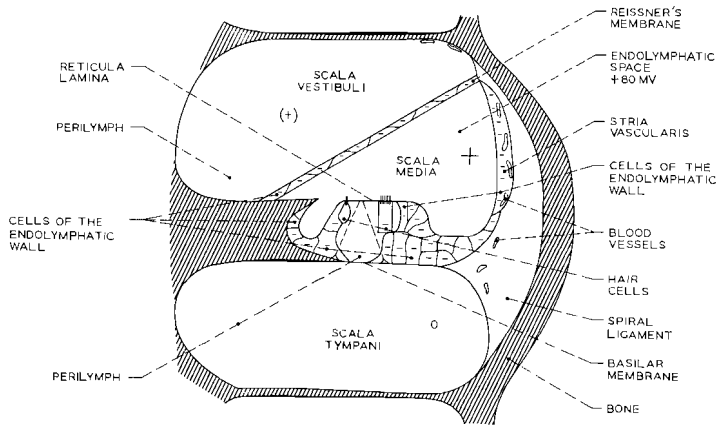

Fig. 4.8. Distribution of resting potentials in the cochlea. Scala tympani is taken as the zero reference. The tectorial membrane is not shown. The interiors of all cells are strongly negative. (After TASAKI, DAVIS and ELDREDGE)

presumably intracellular, are strongly negative. The endolymphatic space (scala media) is strongly positive. (Refer to Fig. 4.5 for more details on the organ of Corti.)

If a microelectrode penetrates upward through the basilar membrane, recording simultaneously the dc potentials (which serve to locate the electrode tip) and the cochlear microphonic response to a 500 cps tone, the result is shown in Fig. 4.9. The conclusion is that the electrical interface at which the phase reversal of the cochlear microphonic occurs is

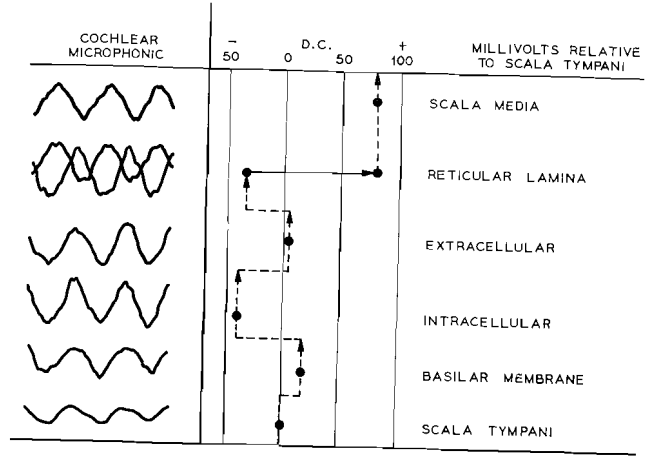

Fig. 4.9. Cochlear microphonic and dc potentials recorded by a microelectrode penetrating the organ of Corti from the scala tympani side. The cochlear microphonic is in response to a 500 cps tone. (After DAVIS, 1965)

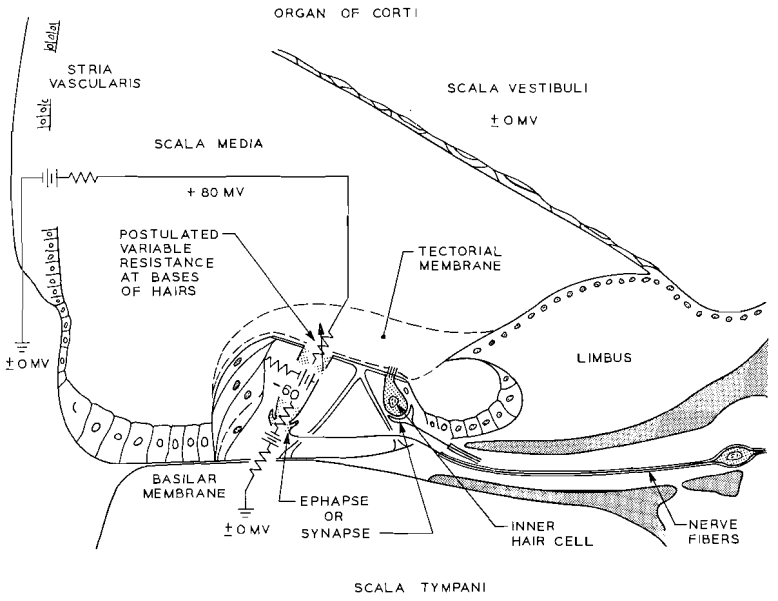

Fig. 4.10. A "resistance microphone" theory of cochlear transduction. (After DAVIS, 1965)

the hair-bearing surface of the hair cell (although one cannot differentiate between the surface and base location of the hair cell).

Two biological batteries therefore act in series: the internal (negative) polarization of the hair cells and the (positive) dc polarization of the endocochlear voltage (which is probably generated by the stria vascularis). This action leads to the conception of the equivalent circuit for cochlear excitation shown in Fig. 4.10.

The cochlear microphonic, as already mentioned, is viewed as the fluctuating voltage drop across the cell membrane due to alternating increase or decrease in its ohmic resistance. It appears to be a facsimile representation of the local displacement of the basilar membrane (TEAS *et al.*). The dynamic range of the microphonic is relatively large. Its amplitude appears linearly related to input sound pressure over ranges greater than 40 dB (TEAS *et al.*).

Although the functional link between the cochlear microphonic (or the facsimile membrane displacement) and the all-or-none electrical activity in the fibers of the auditory nerve remains obscure, it is nevertheless clear that a local deformation of the membrane (of sufficient amplitude), and a consequent bending of the sensory hairs in the area, causes the sensory cells to generate a scala media potential favorable for triggering the neurons in that region. The greater the displacement

magnitude, the greater the number of neurons activated. A periodic displacement of sufficiently low frequency elicits neural firing synchronous with the stimulus. The periodicity of tones of frequencies less than about 1000 cps may therefore be represented by the periodicity of the neural volleys. This mechanism may be one method of coding for the subjective attribute of pitch. The fact that the neurons leading away from a given region of the frequency-selective basilar membrane maintain their identity in the auditory nerve offers a further possibility for the coding of pitch, namely, in terms of membrane place of maximum stimulation.

4.15. Neural Pathways in the Auditory System

A schematic representation of the ascending neural pathways associated with one ear are shown in Fig. 4.11. Beginning in the organ of Corti, the roughly 30000 individual neurons innervate singly or multiply about the same number of sensory (hair) cells. (In general, the inner

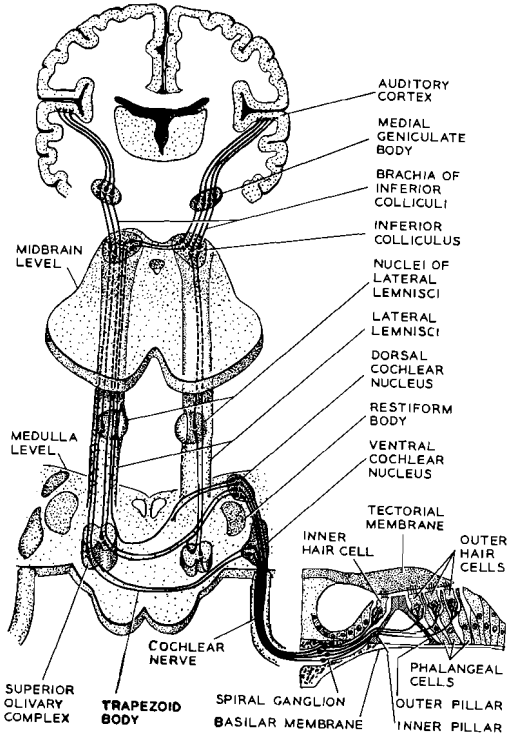

Fig. 4.11. Schematic diagram of the ascending auditory pathways. (Adapted from drawings by NETTER)

hair cells are served by only one or two neurons, the outer cells by several.) The dendritic arbors of the first-order neurons bear on the sensory cells. The cell bodies of the first-order neurons are located in the spiral ganglion, and their axons pass via the cochlear nerve (about 5 mm) to the dorsal and ventral cochlear nuclei in the medulla. Here the first synapses of the auditory pathway are located. From these nuclei, some second-order neurons pass to the superior olive on the same side, some decussate to the opposite side. Some pass upward to the medial geniculate body, with or without intermediate synapses with other neurons located in the lateral lemnisci and the inferior colliculi. The latter nuclei are located in the midbrain, and a second, smaller pathway of decussation runs between them. Thus, stimuli received at the two ears may interact both at the medulla and midbrain levels. The last stage in the pathway is the auditory cortex. The exact neuro-electrical representation of sound stimuli at these various levels is not well understood, and considerable research effort is presently aimed at studying these processes.

The first-order fibers of the auditory nerve connect to different places along the cochlear partition. Starting at the point (apex) of the cochlea, they are progressively collected in the internal auditory meatus until, at the base, the whole nerve trunk is formed. Because the basilar membrane is a mechanical frequency analyzer, it is not surprising that individual fibers exhibit frequency specificity. Owing to the way in which the fibers are collected and the trunk formed, those fibers which have greatest sensitivity to high frequencies lie on the outer part of the whole nerve, while those more sensitive to low frequencies tend to be found toward the core. This "tonotopic" organization of the auditory system (that is, its place-frequency preserving aspect) seems to be maintained at least to some degree all the way to the cortical level (TUNTURI).

The electrical response of individual fibers is a standard pulse. Characteristically, the pulse exhibits a duration on the order of a millisecond. The activity is statistical in two senses. First, the firing patterns of an individual fiber are not identical in successive repetitions of a given stimulus. Second, individual fibers exhibit spontaneous firing (electrical output) of a random nature. The latter appears to be much the same for all first-order fibers.

Comprehensive investigation of first-order electrical behavior in cats has been carried out by KIANG *et al.* Since the structure of the cochlea and auditory nerve follows the same general plan in all mammals, data from these studies should give insight into the human auditory system.

Typical microelectrode recordings from single primary fibers are illustrated in Fig. 4.12. In this instance, the signal comprises 50 msec tone bursts of a 2.3 kcps frequency. The upper recording is from a fiber

that is maximally sensitive to this frequency, while the lower is from a fiber maximally sensitive to 6.6 kcps. The electrical output of the former is highly correlated with the stimulus, while the electrical output of the latter is not. The nerve response potential is recorded with respect to a convenient reference potential, in this case the head holder for the preparation. A positive voltage is indicated by a downward deflection.

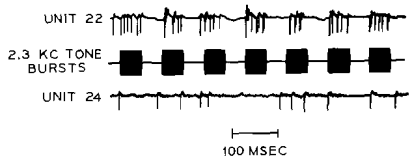

Fig. 4.12. Electrical firings from two auditory nerve fibers. The characteristic frequency of unit 22 is 2.3 kcps and that for unit 24 is 6.6 kcps. The stimulus is 50 msec bursts of a 2.3 kcps tone. (After KIANG *et al.*)

By choosing a suitable criterion of response, the frequency characteristic (tuning curve) of an individual first-order fiber can be measured. Results of such measurements are illustrated for several fibers in Fig. 4.13. The frequency for which the threshold is lowest (the minimum of each curve) is called the characteristic frequency (CF) of the fiber (unit). These minima appear to match well the shape of the audiogram determined from behavioral measurements. An interesting aspect of these data is that while over the low-frequency range the shapes of the tuning curves appear to be nearly constant percentage bandwidth in character (constant Q) and display a bandwidth which correlates reasonably well with BÉKÉSY's mechanical responses, the tuning curves of high-frequency units are much sharper and display Q increasing with frequency. (BÉKÉSY's

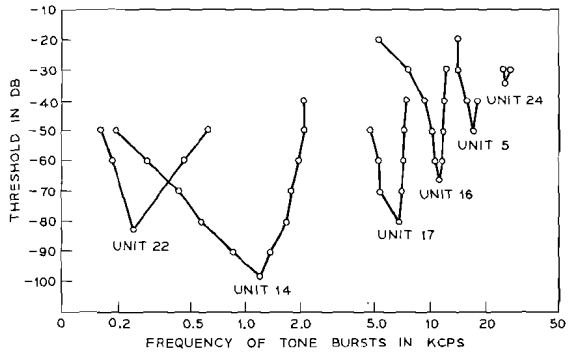

Fig. 4.13. Frequency sensitivities for six different fibers in the auditory nerve of cat. (After KIANG *et al.*)

observations on human basilar membrane were, of course, limited to the low end of the spectrum -2400 cps and down.)

KIANG *et al.* have also observed the electrical response of primary units to punctuate signals, namely, broadband clicks. Individual responses to 10 successive rarefaction clicks of 100-μsec duration are plotted in Fig. 4.14. The figure shows the electrical response recorded at the round window (RW, primarily the cochlear microphonic) and the response of the individual fiber. The time origin is taken as the time the pulse is delivered to the earphone. The characteristic frequency of the unit is 540 cps. The pattern of firing differs in successive stimulations, but the responses show a consistent periodic aspect. Multiple firings in response to a single click are apparent.

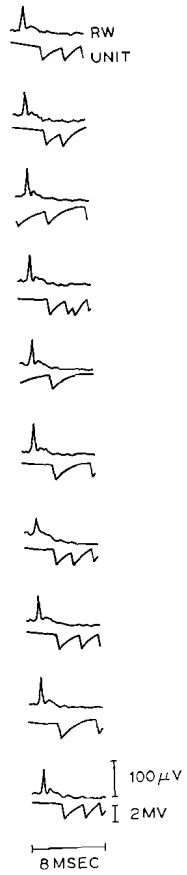

Fig. 4.14. Electrical response of a single auditory nerve fiber (unit) to 10 successive rarefaction pulses of 100 μsec duration. *RW* displays the cochlear microphonic response at the round window. $CF = 540$ cps. (After KIANG *et al.*)

A convenient way to analyze this periodic feature is, in successive presentations of the signal, to measure the number of times the fiber fires at a prescribed time after the signal onset. This number plotted against the time of firing forms the post-stimulus time (PST) histogram. Some quantization of the time scale is implied and this quantization (or "bin width") is made sufficiently small to resolve the periodicities of interest. (For click signals, a bin width of 0.063 msec was customarily used.) A digital computer is a valuable tool for calculating and displaying the histogram. One minute of data from the conditions in Fig. 4.14 produces the histogram of Fig. 4.15. (Since the clicks are

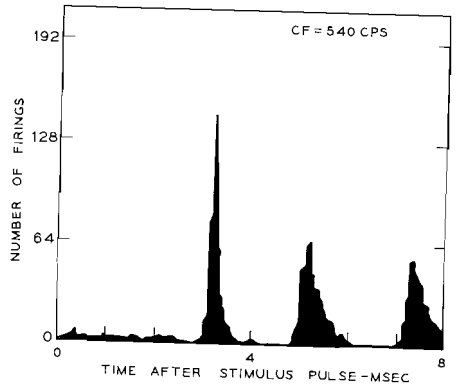

Fig. 4.15. Post stimulus time (PST) histogram for the nerve fiber shown in Fig. 4.14. $CF = 540$ cps. Stimulus pulses 10 sec^{-1}. (After KIANG *et al.*)

delivered at a rate of 10 sec^{-1}, this histogram is the result of 600 signal presentations.) The times of firings show a periodic structure, or "preferred" times for firing. In the midfrequency range for the animal, the histogram may exhibit as many as five or six distinct peaks or preferred times. At the upper end of the animal's frequency range, the tendency is for the histogram to display a single major peak.

The preferred times for firing appear to be intimately linked to the characteristic frequency of the unit, and the interval between peaks in the PST histogram is approximately equal to 1/CF. Higher frequency units consequently show smaller intervals between the PST histogram peaks. The interval between peaks in the histogram and 1/CF are related as shown in Fig. 4.16. Data for 56 different units are plotted. The multiple responses of single primary units to single clicks almost certainly reflect the mechanical response of the cochlea. (See the derivations of Section 4.2 for the impulse response of the basilar membrane.)

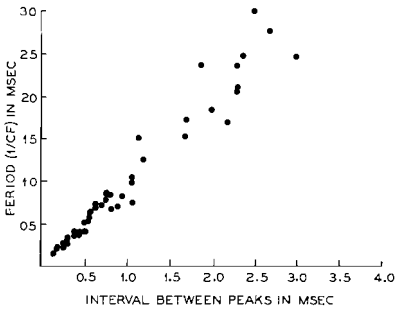

Fig. 4.16. Characteristic period ($1/CF$) for 56 different auditory nerve fibers plotted against the interpeak interval measured from PST histograms. (After KIANG *et al.*)

Microelectrode studies of the electrical activity of single neurons at other levels in the auditory pathway have been, and are being, carried out. Varying experimental techniques and methods of anesthesia have sometimes led to disagreements among the results, but as research progresses the neural schema is becoming increasingly better understood.

According to at least one investigation on cat, the rate of single unit firings is monotonically related to stimulus intensity at all neural stages from the periphery up to the medial geniculate body (KATSUKI). This is exemplified for sinusoidal tones in Figs. 4.17 and 4.18. Fig. 4.17 shows the spikes (firings) of a single neuron in the trapezoidal body of cat in response to tone bursts of 9000 cps, delivered at four different levels. The spike duration is on the order of the conventional 1 msec, and the firings are more numerous for the more intense sounds.

Fig. 4.18 shows a monotone relation between firing rate and intensity for different neural stages. The firing rate for the first-order single neuron (the top curve for the cochlear nerve) has a maximum value close to its best (characteristic) frequency, namely 830 cps. This suggests that for the sinusoidal stimulation, the first-order neuron fires at most once per period. The rates at the higher neural stages appear substantially less than their characteristic frequencies.

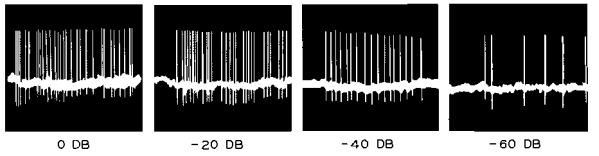

Fig. 4.17. Responses of a single auditory neuron in the trapezoidal body of cat. The stimulus was tone bursts of 9000 cps produced at the indicated relative intensities. (After KATSUKI)

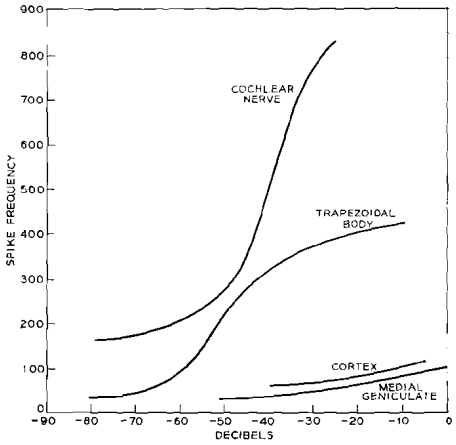

Fig. 4.18. Relation between sound intensity and firing (spike) frequency for single neurons at four different neural stages in the auditory tract of cat. Characteristic frequencies of the single units: Nerve: 830 cps; Trapezoid: 9000 cps; Cortex: 3500 cps; Geniculate: 6000 cps. (After KATSUKI)

Microelectrode recordings from single first-order neurons often show appreciable spontaneous activity. At higher neural stages and in the cortex, spontaneous activity apparently is not as pronounced (KATSUKI).

The cochlear nucleus complex of cat has been another particular area of study (ROSE, GALAMBOS and HUGHES). Strong evidence for a distinct tonotopical organization is found in the major subdivision of the cochlear nucleus. Typical of this finding is the sagittal section through the left cochlear complex shown in Fig. 4.19. The frequency scale indicates the best (most sensitive) frequencies of the neurons located along the ruled axis.

Some tonotopical organization appears to exist at the cortical level, although its degree and extent seems to be controversial (for example, KATSUKI; TUNTURI).

The relations between threshold sound amplitude and tone frequency (that is, the tuning curves) for single units at the cochlear nucleus level have been found to vary in shape (ROSE, GALAMBOS and HUGHES). Some appear broad, others narrow. All, however, resemble roughly the mechanical resonance characteristic of the basilar membrane. That is, the tuning curve (or threshold amplitude) rises more steeply on the high-frequency side than on the low-frequency side. Typical narrow and broad tuning curves obtained from single units in the cochlear nucleus are shown in Fig. 4.20a and b, respectively. For tones up to about 60 db above the threshold, the frequency range of response for both narrow and broad units does

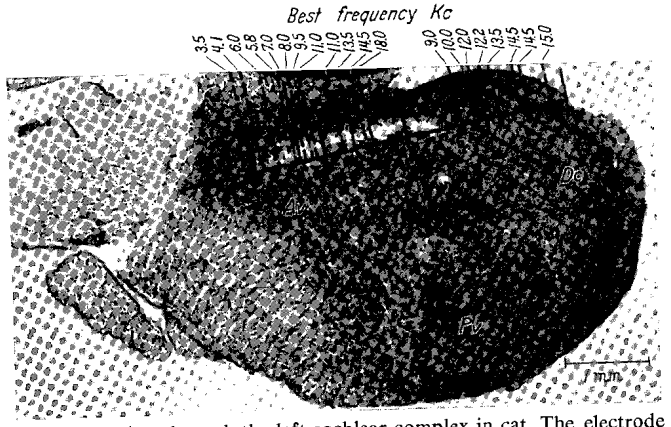

Fig. 4.19. Sagittal section through the left cochlear complex in cat. The electrode followed the track visible just above the ruled line. Frequencies of best response of neurons along the track are indicated. (After ROSE, GALAMBOS and HUGHES)

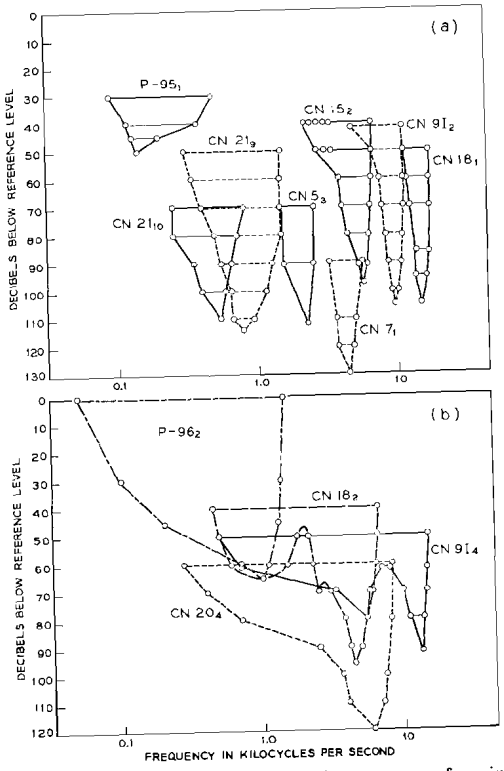

Fig. 4.20a and b. Intensity *vs* frequency "threshold" responses for single neurons in the cochlear nucleus of cat. The different curves represent the responses of different neurons. (a) Units with narrow response areas; (b) units with broad response areas. (After ROSE, GALAMBOS and HUGHES)

not extend over more than about 0.3 of an octave above the best frequency. The frequency range below the best frequency can range from about 0.4 to 3.8 octaves for the narrow units, to almost the whole lower frequency range for the broad units. Single units at this level display adaptive and inhibitory behavior which is strongly intensity dependent.

The mechanism of neural transmission across a synapse also remains to be firmly established. A temporal delay – typically on the order of 1 msec – is usually incurred at the junction. Response latencies at the level of the cochlear nucleus have minimum times on the order of 2 to 3 msec, but latencies as great as 6 to 8 msec have been measured. At the cortical level, latencies as great as 20 to 30 msec and as small as 6 to 8 msec are possible.

4.2. Computational Models for Ear Function

It has been emphasized in the preceding discussion that the complete mechanism of auditory perception is far from being adequately understood. Even so, present knowledge of ear physiology, nerve electrophysiology, and subjective behavior make it possible to relate certain auditory functions among these disparate realms. Such correlations are facilitated if behavior can be quantified and analytically specified. As a step in this direction, a computational model has been derived to describe basilar membrane displacement in response to an arbitrary sound pressure at the eardrum (FLANAGAN, 1962a).

The physiological functions embraced by the model are shown in the upper diagram of Fig. 4.21. In this simplified schematic of the peripheral

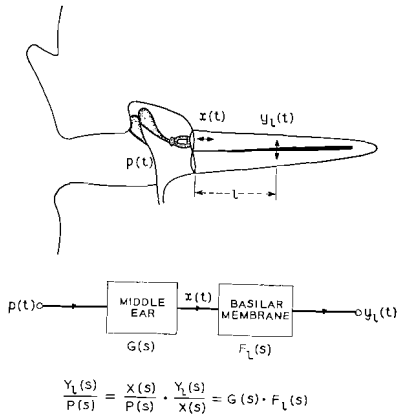

Fig. 4.21. Schematic diagram of the peripheral ear. The quantities to be related analytically are the eardrum pressure, $p(t)$; the stapes displacement, $x(t)$; and the basilar membrane displacement at distance l from the stapes, $y_l(t)$

ear, the cochlea is shown uncoiled. $p(t)$ is the sound pressure at the eardrum, $x(t)$ is the equivalent linear displacement of the stapes footplate, and $y_l(t)$ is the linear displacement of the basilar membrane at a distance l from the stapes. The desired objective is an analytical approximation to the relations among these quantities. It is convenient to obtain it in two steps. The first step is to approximate the middle-ear transmission, that is, the relation between $x(t)$ and $p(t)$. The second is to approximate the transmission from the stapes to the specified point l on the membrane. The approximating functions are indicated as the frequency-domain (LAPLACE) transforms $G(s)$ and $F_l(s)$, respectively, in the lower part of Fig. 4.21.

The functions $G(s)$ and $F_l(s)$ must be fitted to available physiological data. If the ear is assumed to be mechanically passive and linear over the frequency and amplitude ranges of interest, rational functions of frequency with stable normal modes (left half-plane poles) can be used to approximate the physiological data. Besides computational convenience, the rational functions have the advantage that they can be realized in terms of lumped-constant electrical circuits, if desired. Because the model is an input-output or "terminal" analog, the response of one point does not require explicit computation of the activity at other points. One therefore has the freedom to calculate the displacement $y_l(t)$ for as many, or for as few, values of l as are desired.

4.21. Basilar Membrane Model

The physiological data upon which the form of $F_l(s)$ is based are those of BÉKÉSY, shown in Fig. 4.6[1]. If the curves of Fig. 4.6 are normalized with respect to the frequency of the maximum response, one finds that they are approximately constant percentage bandwidth responses. One also finds that the phase data suggest a component which is approximately a simple delay, and whose value is inversely proportional to the frequency of peak response. That is, low frequency points on the membrane (nearer the apex) exhibit more delay than high frequency (basal) points. A more detailed discussion of these relations and the functional fitting of the data has been given previously (FLANAGAN, 1962a). [In this earlier work, the fit afforded by three different forms of $F_l(s)$ was considered. For purpose of the present discussion, only the results for the first, a fifth-degree function, will be used.]

The physiological data can, of course, be approximated as closely as desired by selecting an appropriately complex model. The present model

[1] More recent data on basilar membrane vibration, determined in animal experiments using the Mössbauer effect (JOHNSTONE and BOYLE; RHODE), may also serve as this point of departure.

is chosen to be a realistic compromise between computational tractability and adequacy in specifying the physiological data. One function which provides a reasonable fit to BÉKÉSY's results is

$$F_l(s)=c_1\beta_l^4\left(\frac{2000\pi\beta_l}{\beta_l+2000\pi}\right)^{0.8}\left(\frac{s+\varepsilon_l}{s+\beta_l}\right)\left[\frac{1}{(s+\alpha_l)^2+\beta_l^2}\right]^2 e^{\frac{-3\pi s}{4\beta_l}}, \quad (4.1)$$

where

$s=\sigma+j\omega$ is the complex frequency,

$\beta_l=2\alpha_l$ is the radian frequency to which the point l-distance from the stapes responds maximally,

c_1 is a real constant that gives the proper absolute value of displacement,

$e^{\frac{-3\pi s}{4\beta_l}}$ is a delay factor of $3\pi/4\beta_l$ seconds which brings the phase delay of the model into line with the phase measured on the human ear. This factor is primarily transit delay from stapes to point l on the membrane,

$\left(\frac{2000\pi\beta_l}{\beta_l+2000\pi}\right)^{0.8}\beta_l^4$ is an amplitude factor which matches the variations in peak response with resonant frequency β_l, as measured physiologically by BÉKÉSY (1943).

$\varepsilon_l/\beta_l=0.1$ to 0.0 depending upon the desired fit to the response at low frequencies.

The membrane response at any point is therefore approximated in terms of the poles and zeros of the rational function part of $F_l(s)$. As indicated previously in Fig. 4.6, the resonant properties of the membrane are approximately constant-Q (constant percentage bandwidth) in character. The real and imaginary parts of the critical frequencies can therefore be related by a constant factor, namely, $\beta_l=2\alpha_l$. To within a multiplicative constant, then, the imaginary part of the pole frequency, β_l, completely describes the model and the characteristics of the membrane at a place l-distance from the stapes. The pole-zero diagram for the model is shown in Fig. 4.22a.

The real-frequency response of the model is evidenced by letting $s=j\omega$. If frequency is normalized in terms of $\zeta=\omega/\beta_l$, then relative phase and amplitude responses of $F_l(j\zeta)$ are as shown in Fig. 4.22b. Because of the previously mentioned relations, $F_l(\zeta)$ has (except for the multiplicative constant) the same form for all values of l.

The inverse Laplace transform of (4.1) is the displacement response of the membrane to an impulse of displacement by the stapes. The details of the inverse transformation are numerically lengthy, but if the

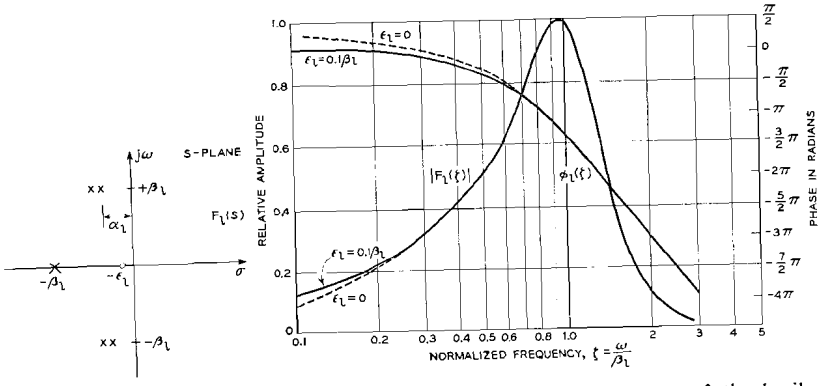

Fig. 4.22a. Pole-zero diagram for the approximating function $F_l(s)$. (After FLANAGAN, 1962a)

Fig. 4.22b. Amplitude and phase response of the basilar membrane model $F_l(s)$. Frequency is normalized in terms of the characteristic frequency β_l

mathematics is followed through it is found to be

$$\begin{aligned} f_l(t)=c_1\left(\frac{2000\pi}{\beta_l+2000\pi}\right)^{0.8}\beta_l^{1+r}\{&[0.033+0.360\beta_l(t-T)]\\ &\times e^{-\frac{\beta_l(t-T)}{2}}\sin\beta_l(t-T)+[0.575-0.320\beta_l(t-T)]\\ &\times e^{-\frac{\beta_l(t-T)}{2}}\cos\beta_l(t-T)-0.575\,e^{-\beta_l(t-T)}\}=0 \end{aligned} \quad (4.2)$$

$$\text{for } t\geqq T \quad \text{and} \quad \varepsilon_l/\beta_l=0.1,$$

where the delay $T=3\pi/4\beta_l$, as previously stated. A plot of the response (4.2) is shown in Fig. 4.23.

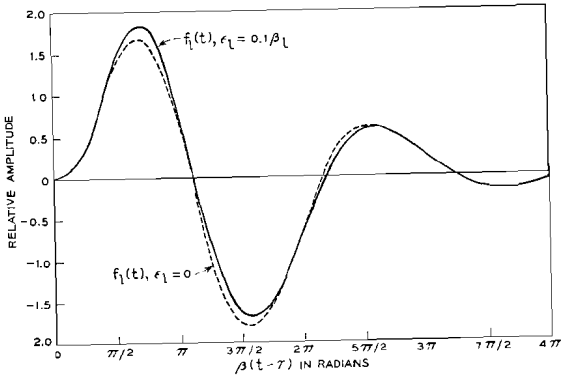

Fig. 4.23. Response of the basilar membrane model to an impulse of stapesdisplacement

Note, too, from the form of (4.1) that the complex displacement response can be determined as a function of the place frequency β_l for a given stimulating frequency $s=j\omega_n$. The radian frequency β_l can, in turn, be related directly to the distance l (in mm) from the stapes by

$$(35-l)=7.5\log\beta_l/2\pi(20)$$

(see FLANAGAN, 1962a). Therefore (4.1) can be used to compute $F(s,l)|_{s=j\omega_n}=A(l)\,e^{j\varphi(l)}$ to give spatial responses of amplitude and phase similar to those shown in Fig. 4.6c.

4.22. Middle Ear Transmission

To account for middle ear transmission, an analytical specification is necessary of the stapes displacement produced by a given sound pressure at the eardrum (see Fig. 4.21). Quantitative physioacoustical data on the operation of the human middle ear are sparse. The data which are available are due largely to BÉKÉSY and, later, to ZWISLOCKI and to MØLLER. These results have been shown in Fig. 4.3. The data suggest appreciable variability and uncertainty, particularly in connection with the critical (roll-off) frequency and damping of the characteristic. All agree, however, that the middle ear transmission is a low-pass function. BÉKÉSY's results were obtained from physiological measurements. ZWISLOCKI's and MØLLER's data are from electrical analogs based upon impedance measurements at the eardrum, a knowledge of the topology of the middle ear circuit, and a knowledge of some of the circuit constants. In gross respects the data are in agreement[1].

If ZWISLOCKI's results in Fig. 4.3 are used, they can be approximated reasonably well by a function of third degree. Such an approximating function is of the form

$$G(s)=\frac{c_0}{(s+a)\,[(s+a)^2+b^2]}, \tag{4.3}$$

where c_0 is a positive real constant. [When combined with $F_l(s)$, the multiplying constants are chosen to yield proper absolute membrane displacement. For convenience, one might consider $c_0=a(a^2+b^2)$ so that the low-frequency transmission of $G(s)$ is unity.] When the pole frequencies of $G(s)$ are related according to

$$b=2a=2\pi(1500)\,\text{rad/sec}, \tag{4.4}$$

the fit to ZWISLOCKI's data is shown by the plotted points in Fig. 4.24.

The inverse transform of (4.3) is the displacement response of the stapes to an impulse of pressure at the eardrum. It is easily obtained and

[1] Recent measurements on middle-ear transmission in cat (GUINAN and PEAKE) also correspond favorably with these data.

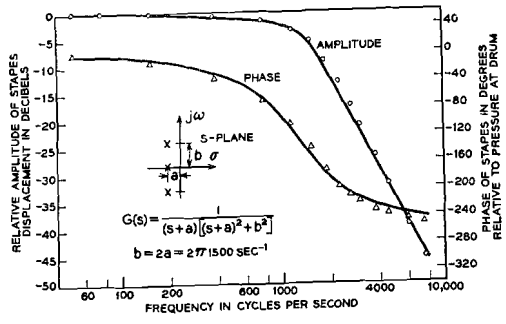

Fig. 4.24. Functional approximation of middle ear transmission. The solid curves are from an electrical analog by ZWISLOCKI (see Fig. 4.3c). The plotted points are amplitude and phase values of the approximating function $G(s)$. (FLANAGAN, 1962a)

will be useful in the subsequent discussion. Let

$$G(s)=G_1(s)\,G_2(s),$$

where

$$G_1(s)=\frac{c_0}{s+a}; \qquad G_2(s)=\frac{1}{(s+a)^2+b^2}. \tag{4.5}$$

The inverses of the parts are

$$g_1(t)=c_0\,e^{-at}; \qquad g_2(t)=\frac{e^{-at}}{b}\sin bt. \tag{4.6}$$

The inverse of $G(s)$ is then the convolution of $g_1(t)$ and $g_2(t)$

$$g(t)=\int_0^t g_1(\tau)\,g_2(t-\tau)\,d\tau,$$

or

$$g(t)=c_0\frac{e^{-at}}{b}(1-\cos bt)=\frac{c_0\,e^{-bt/2}}{b}(1-\cos bt). \tag{4.7}$$

Also for future use, note that the time derivative of the stapes displacement is

$$\dot{g}(t)=\frac{c_0\,e^{-bt/2}}{2}(2\sin bt+\cos bt-1). \tag{4.8}$$

Plots of $g(t)$ and $\dot{g}(t)$ are shown in Fig. 4.25. For this middle ear function, the response is seen to be heavily damped. Other data, for example MØLLER's in Fig. 4.3, suggest somewhat less damping and the possibility of adequate approximation by a still simpler, second-degree

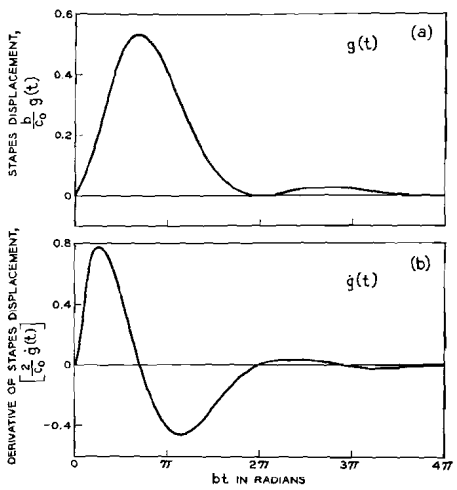

Fig. 4.25a and b. Displacement and velocity responses of the stapes to an impulse of pressure at the eardrum

function. For such a transmission, the stapes impulse response would be somewhat more oscillatory[1].

4.23. Combined Response of Middle Ear and Basilar Membrane

The combined response of the models for the middle ear and basilar membrane is

$$H_l(s) = G(s)\,F_l(s)$$
$$h_l(t) = g(t) * f_l(t). \tag{4.9}$$

For the $F_l(s)$ model described here, the combined time response is easiest obtained by inverse transforming $H_l(s)$. [For other $F_l(s)$ models, the combined response may be more conveniently computed from time-domain convolution.]

The details of the inverse transform of $H_l(s)$ are numerically involved and only the result is of interest here. When the inverse transform is calculated, the result has the form

$$\begin{aligned} h_l(\tau) = {} & A e^{-b\tau/2} + B e^{-b\tau/2}(\cos b\tau - \tfrac{1}{2}\sin b\tau) + C(e^{-b\tau/2}\sin b\tau) \\ & + D e^{-\eta b\tau} + E(e^{-\eta b\tau/2}\sin \eta b\tau) + F(\eta b\tau\, e^{-\eta b\tau/2}\sin b\tau) \\ & + G(e^{-\eta b\tau/2}\cos \eta b\tau) + H(\eta b\tau\, e^{-\eta b\tau/2}\cos \eta b\tau); \quad \text{for } \tau \geqq 0, \end{aligned} \tag{4.10}$$

[1] The modelling technique does not of course depend critically upon the particular set of data being modeled. When more complete physiological measurements are forthcoming, the rational function can be altered to fit the new data.

where $\tau=(t-T)$; $T=3\pi/4\beta_l$; $\eta=\beta_l/b$; $\beta_l=2\alpha_l$; $b=2a$; $\varepsilon_l=0$; and the A, B, C, D, E, F, G, H are all real numbers which are functions of β_l and b (see FLANAGAN, 1962a, for explicit description).

The form of the impulse response is thus seen to depend upon the parameter $\eta=\beta_l/b$. Values of $\eta<1.0$ refer to (apical) membrane points whose frequency of maximal response is less than the critical frequency of the middle ear. For these points, the middle-ear transmission is essentially constant with frequency, and the membrane displacement is very nearly that indicated by $f_l(t)$ in Eq. (4.2). On the other hand, values of $\eta>1.0$ refer to (basal) points which respond maximally at frequencies greater than the critical frequency of the middle ear. For these points, the middle-ear transmission is highly dependent upon frequency and would be expected to influence strongly the membrane displacement. To illustrate this point, Eq. (4.10) has been evaluated for $\eta=0.1$, 0.8, and 3.0. The result is shown in Fig. 4.26.

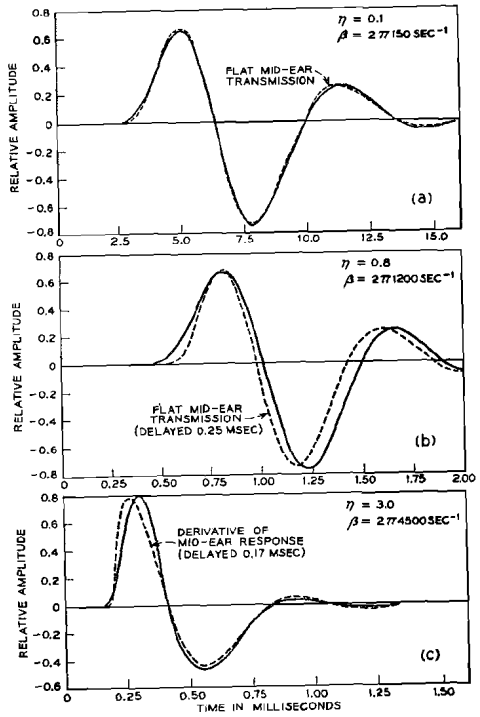

Fig. 4.26a–c. Displacement responses for apical, middle and basal points on the membrane to an impulse of pressure at the eardrum. The responses are computed from the inverse transform of $[G(s)\,F_l(s)]$

For an impulse of pressure delivered to the eardrum, the three solid curves represent the membrane displacements at points which respond maximally to frequencies of 150, 1200, and 4500 cps, respectively. Each of the plots also includes a dashed curve. In Figs. 4.26a and 4.26b, the dashed curve is the membrane displacement computed by assuming the middle-ear transmission to be constant, or flat, and with zero phase. This is simply the response $[\mathscr{L}^{-1} F_l(s)]$. In Fig. 4.26c the dashed curve is the time derivative of the stapes displacement, $g(t)$, taken from Fig. 4.25. Fig. 4.25c therefore suggests that the form of the membrane displacement in the basal region is very similar to the derivative of the stapes displacement.

The individual frequency-domain responses for $G(s)$ and $F_l(s)$ have been shown in Figs. 4.22 and 4.24, respectively. The combined response in the frequency domain is simply the sum of the individual curves for amplitude (in db) and phase (in radians). The combined amplitude and phase responses for the model $G(s)F_l(s)$ are shown in Figs. 4.27a and 4.27b, respectively.

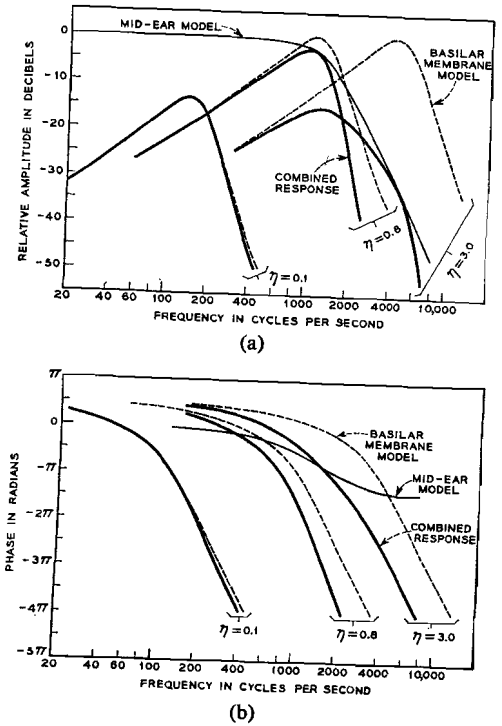

Fig. 4.27a and b. (a) Amplitude *vs* frequency responses for the combined model. (b) Phase *vs* frequency responses for the combined model

As already indicated by the impulse responses, the response of apical (low-frequency) points on the membrane is given essentially by $F_l(s)$, while for basal (high-frequency) points the response is considerably influenced by the middle-ear transmission $G(s)$. Concerning the latter point, two things may be noted about the frequency response of the membrane model [i.e., $F_l(\omega)$]. First, the low-frequency skirt of the amplitude curve rises at about 6 db/octave. And second, the phase of the membrane model [i.e. $\underline{/F_l(\omega)}$] approaches $+\pi/2$ radians at frequencies below the peak amplitude response[1]. In other words, at frequencies appreciably less than its peak response frequency, the membrane function $F_l(\omega)$ behaves crudely as a differentiator. Because the middle-ear transmission begins to diminish in amplitude at frequencies above about 1500 cps, the membrane displacement in the basal region is roughly the time derivative of the stapes displacement. The waveform of the impulse response along the basal part of the membrane is therefore approximately constant in shape. Along the apical part, however, the impulse response oscillates more slowly (in time) as the apex is approached. This has already been illustrated in Fig. 4.26.

One further point may be noted from Fig. 4.27. Because the amplitude response of the middle-ear declines appreciably at high frequencies, the amplitude response of a basal point is highly asymmetrical. (Note the combined response for $\eta=3.0$.) The result is that a given basal point—while responding with greater amplitude than any other membrane point at its characteristic frequency—responds with greatest amplitude (but not greater than some other point) at some lower frequency.

4.24. An Electrical Circuit for Simulating Basilar Membrane Displacement

On the basis of the relations developed in the previous sections [Eqs. (4.1) and (4.3)], it is possible to construct electrical circuits whose transmission properties are identical to those of the functions $G(s)$ and $F_l(s)$. This is easiest done by representing the critical frequencies in terms of simple cascaded resonant circuits, and supplying the additional phase delay by means of an electrical delay line. Such a simulation for the condition $\varepsilon_l=0$ is shown in Fig. 4.28.

The voltage at an individual output tap represents the membrane displacement at a specified distance from the stapes. The electrical

[1] This phase behavior is contrary to the physiological phase measurements shown in Fig. 4.6b. Nevertheless, calculations of minimum phase responses for the basilar membrane indicated that the low-frequency phase behavior must approach $\pi/2$ radians *lead* (FLANAGAN and BIRD). This earlier analytical prediction (and hence justification for the choice $\varepsilon_l=0$) has been confirmed by recent measurements. These measurements, using the Mössbauer effect, in fact reveal a leading phase at low frequencies (JOHNSTONE and BOYLE; RHODE).

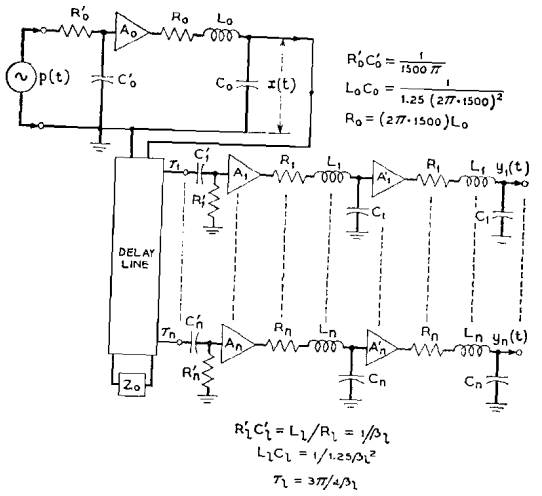

Fig. 4.28. Electrical network representation of the ear model

voltages analogous to the sound pressure at the eardrum and to the stapes displacement are also indicated. The buffer amplifiers labelled A have fixed gains which take account of the proper multiplicative amplitude constants.

The circuit elements are selected according to the constraints stated for $G(s)$ and $F_l(s)$. The constraints are represented by the equations shown in Fig. 4.28 and, together with choice of impedance levels, completely specify the circuit. For each membrane point the relative gains of the amplifiers are set to satisfy the amplitude relations implied in Fig. 4.27a. The gains also take account of the constant multiplying factors in the rational function models.

Some representative impulse responses of the analog circuit of Fig. 4.28 are shown in Fig. 4.29a. One notices the degradation in time resolution as the response is viewed at points more apicalward. That is, the frequency resolution of the membrane increases as the apex is approached.

The electrical circuit can also be used in a simple manner to provide an approximation to the spatial derivative of displacement. This function, like the displacement, may be important in the conversion of mechanical-to-neural activity. As mentioned earlier, it has been noted that the inner hair cells in the organ of Corti appear sensitive to longitudinal bending of the membrane, whereas the outer cells are sensitive to transverse bending (BÉKÉSY, 1953). The former may therefore be more sensitive to the spatial gradient or derivative of membrane displacement, while the latter may be primarily sensitive to displacement.

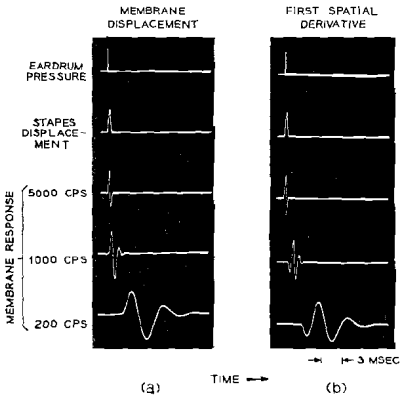

Fig. 4.29a and b. (a) Impulse responses measured on the network of Fig. 4.28. (b) First difference approximations to the spatial derivative measured from the network of Fig. 4.28

The differences between the deflection of adjacent, uniformly-spaced points can be taken as an approximation to the spatial derivative. Fig. 4.29b shows the first spatial difference obtained from the analog circuit by taking

$$\frac{\partial y}{\partial x} = \frac{y(t, x+\Delta x) - y(t, x)}{\Delta x},$$

where

$$\Delta x = 0.3 \text{ mm}.$$

The similarity to the displacement is considerable.

4.25. Computer Simulation of Membrane Motion

If it is desired to simulate the membrane motion at a large number of points and to perform complex operations upon the displacement responses, it is convenient to have a digital representation of the model suitable for calculations in a digital computer. One such digital simulation represents the membrane motion at 40 points (FLANAGAN, 1962b).

As might be done in realizing the analog electrical circuit, the digital representation of the model can be constructed from sampled-data equivalents of the individual complex pole-pairs and the individual real poles and zeros. The sampled-data equivalents approximate the continuous functions over the frequency range of interest. The computer operations used to simulate the necessary poles and zeros are shown in Fig. 4.30. All of the square boxes labelled D are delays equal to the time between successive digital samples. The input sampling frequency, $1/D$, in the present simulation is 20 Kcps, and the input data is quantized

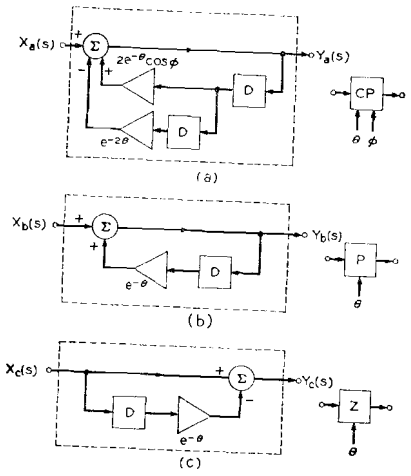

Fig. 4.30. Sampled-data equivalents for the conjugate complex poles, real-axis pole, and real-axis zero

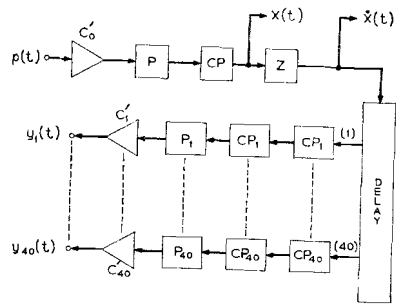

Fig. 4.31. Functional block diagram for a digital computer simulation of basilar membrane displacement

to 11 bits. All of the triangular boxes are "amplifiers" which multiply their input samples by the gain factors shown next to the boxes.

Each of the digital operations enclosed by dashed lines is treated as a component block in the program. The block shown in Fig. 4.30a is labelled CP for conjugate-pole. It has the transfer function

$$\frac{Y_a(s)}{X_a(s)}=[e^{-2\vartheta}e^{-2sD}-2e^{-\vartheta}\cos\Phi e^{-sD}+1]^{-1} \tag{4.11}$$

which has poles at

$$e^{-(\vartheta+sD)}=\cos\Phi\pm j\sin\Phi,$$

or

$$s=\frac{1}{D}[-\vartheta\pm j(\Phi+2n\pi)], \quad n=0, 1, 2, \ldots,$$

so that

$$\vartheta_l=\alpha_l D \quad \text{and} \quad \Phi_l=\beta_l D,$$

where α_l and β_l are the real and imaginary parts of the pole-pair to be simulated The pole constellation of the sampled-data function repeats at $\pm j2n\pi/D$ (or at $\pm j2n\pi/5\times10^{-5}$ for the 20 kcps sampling frequency).

Single real-axis poles are approximated as shown by the P block in Fig. 4.30b. The transfer function is

$$\frac{Y_b(s)}{X_b(s)}=[1-e^{-(\vartheta+sD)}]^{-1} \tag{4.12}$$

and has poles at

$$s=\frac{1}{D}(-\vartheta\pm j2n\pi), \quad n=0, 1, 2, \ldots.$$

The single zero is simulated by the Z block in Fig. 4.30c. Its transfer function is the reciprocal of the P block and is

$$\frac{Y_c(s)}{X_c(s)}=1-e^{-(\vartheta+sD)} \tag{4.13}$$

with zeros at

$$s=\frac{1}{D}(-\vartheta\pm j2n\pi), \quad n=0, 1, 2, \ldots.$$

In the present simulation the zero is placed at the origin, so that $\vartheta=0$ (i.e., $\varepsilon_l=0$).

The computer operations diagrammed by these blocks were used to simulate the model $G(s)\,F_l(s)$ for 40 points along the basilar membrane. The points represent 0.5 mm increments in distance along the membrane, and they span the frequency range 75 to 4600 cps. The blocks are put together in the computer program as shown in Fig. 4.31[1]. The amplifier boxes c_0' and c_l' in Fig. 4.31 take into account not only the model amplitude constants c_0 and c_1 and the $(2000\pi\beta_l/\beta_l+2000\pi)^{0.8}$ factor, but also the amplitude responses of the digital component blocks. For example, it is convenient to make the zero-frequency gain of the CP boxes unity, so each c_l' amplifier effectively includes a $[e^{-2\vartheta}-2e^{-\vartheta}\times\cos\Phi+1]^2$ term. The overall effect of the c_0' and c_l' gain adjustments is to yield the amplitudes specified by $G(s)\,F_l(s)$. The delay to each membrane point, $3\pi/4\beta_l$, is simulated in terms of integral numbers of sample intervals. In the present simulation it is consequently represented to the nearest 50 μsec.

An illustrative impulse response from the simulation, plotted automatically by the computer, is shown in Fig. 4.32. The displacement response of the membrane at 40 points is shown as a function of time. The characteristic frequencies of the membrane points are marked along the y-axis, starting with 4600 cps at the lower (basal) end and going to 75 cps at the upper (apical) end. Time is represented along the x-axis. The input pressure signal $p(t)$ is a single positive pulse 100 μsec in duration and delivered at $t=0$. The responses show that the basal points respond with short latency and preserve a relatively broad-band version of the input pulse. The apical points display increasingly greater latency and progressive elimination of high-frequency content from the signal.

[1] In the present case the simulation was facilitated by casting the operations in the format of a special compiler program (see KELLY, VYSSOTSKY and LOCHBAUM).

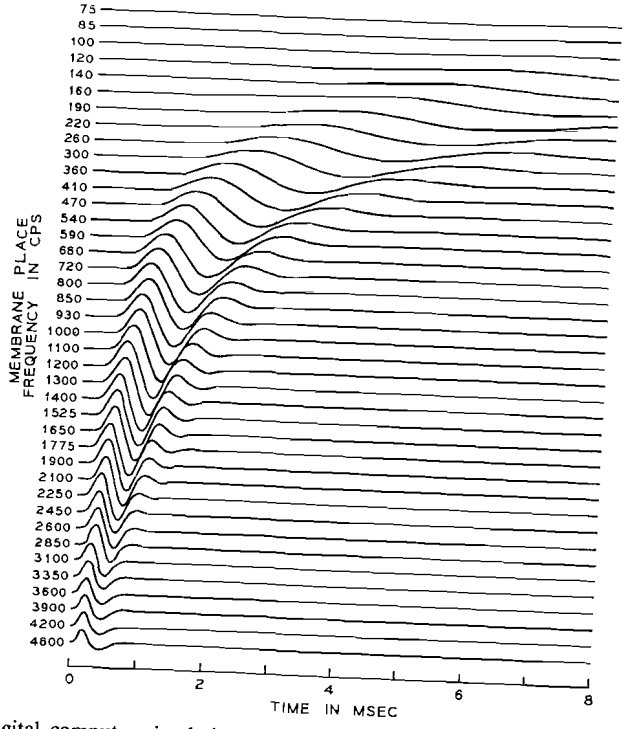

Fig. 4.32. Digital computer simulation of the impulse responses for 40 points along the basilar membrane. The input signal is a single rarefaction pulse, 100 μsec in duration, delivered to the eardrum at time $t = 0$. (After FLANAGAN, 1962b)

These same attributes of the membrane are put in evidence by a periodic pulse signal, which will be of interest in the subsequent discussion. Fig. 4.33 shows the reponse to an input signal composed of alternate positive and negative pulses of 100 μsec duration, produced at a fundamental frequency of 100 cps and initiated at $t=0$. The time between alternate pulses is therefore 5 msec. At the apical (low-frequency) end of the membrane, the frequency resolution is best, and the displacement builds up to the fundamental sinusoid. At the basal (high-frequency) end, the membrane resolves the individual pulses in time. The responses also reflect the transit delay along the membrane.

The utility of the computation model depends equally upon its mathematical tractability and its adequacy in approximating membrane characteristics. Given both, the model can find direct application in relating subjective and physiological auditory behavior. More specifically, it can be useful in relating psychoacoustic responses to patterns of membrane displacement and in establishing an explanatory framework for the neural representation of auditory information.

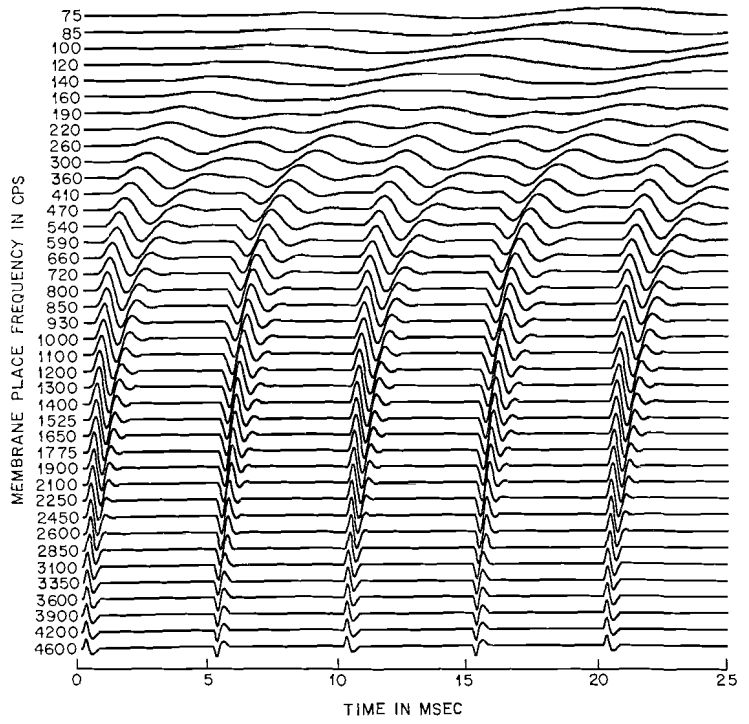

Fig. 4.33. Digital computer output for 40 simulated points along the basilar membrane. Each trace is the displacement response of a given membrane place to alternate positive and negative pressure pulses. The pulses have 100 μsec duration and are produced at a rate of 200 sec^{-1}. The input signal is applied at the eardrum and is initiated at time zero The simulated membrane points are spaced by 0.5 mm. Their characteristic frequencies are indicated along the ordinate. (After FLANAGAN, 1962b)

4.26. Transmission Line Analogs of the Cochlea

The preceding discussion has concerned an "input-output" formulation of the properties of the middle ear and basilar membrane. This approach, for computational and applicational convenience, treats the mechanism in terms of its terminal characteristics. A number of derivations have been made, however, in which the distributed nature of the inner ear is taken into account, and the detailed functioning of the mechanism is examined (PETERSON and BOGERT; BOGERT, 1951; RANKE; ZWISLOCKI, 1948; OETINGER and HAUSER). At least two of these treatments have yielded transmission line analogs for the inner ear.

The simplifying assumptions made in formulating the several treatments are somewhat similar. By way of illustration, they will be indicated for one formulation (PETERSON and BOGERT). The cochlea is idealized

as shown in Fig. 4.34. The oval window is located at O and the round window at R. The distance along the cochlea is reckoned from the base and denoted as x. The cross-sectional areas of the scalas vestibuli and tympani are assumed to be identical functions of distance, $S_0(x)$. The width of the basilar membrane is taken as $b(x)$, and the per-unit-area distributed mass, resistance and stiffness of the basilar membrane (or, more precisely, of the cochlear duct separating the scalas) are respectively $m(x)$, $r(x)$ and $k(x)$. The mechanical constants used are deduced from the physiological measurements of BÉKÉSY.

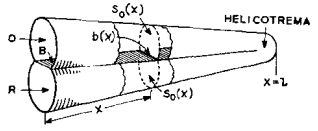

Fig. 4.34. Idealized schematic of the cochlea. (After PETERSON and BOGERT)

The following simplifying assumptions are made. All amplitude are small enough that non-linear effects are excluded. The stapes produces only plane compressional waves in the scalas. Linear relations exists between the pressure difference across the membrane at any point and the membrane displacement, velocity and acceleration at that point. The vertical component of particle velocity in the perilymph fluid is small and is neglected. A given differential element of the membrane exerts no mutual mechanical coupling on its adjacent elements.

The relations necessary to describe the system are the equations for a plane compressional wave propagating in the scalas and the equation of motion for a given membrane element. For a plane wave in the scalas, the sound pressure, p, and particle velocity, u, are linked by the equation of motion

$$\rho \frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x}, \tag{4.14}$$

where ρ is the average density of the perilymph fluid. If the membrane displacements are small, the equations of continuity (mass conservation) for the two scalas are

$$\begin{aligned} \frac{\partial (u_v S)}{\partial x} &= -\frac{S}{\rho c^2}\frac{\partial p_v}{\partial t} - v b \\ \frac{\partial (u_t S)}{\partial x} &= -\frac{S}{\rho c^2}\frac{\partial p_t}{\partial t} + v b, \end{aligned} \tag{4.15}$$

where v is the membrane velocity and the subscripts v and t denote vestibuli and tympani, respectively. These relations state that the rate of mass accumulation for an elemental volume in the scala is equal to the temporal derivative of the fluid density.

The equation of motion for the membrane is

$$(p_v - p_t) = m\frac{dv}{dt} + rv + k\int v\,dt, \tag{4.16}$$

where the pressure difference between the scalas $(p_v - p_t)$ is the forcing function for a membrane element.

Eqs. (4.14) to (4.16) can be solved simultaneously for the pressures and velocities involved. A typical solution for the instantaneous pressure difference produced across the membrane by an excitation of 1000 cps is shown in Fig. 4.35. The pressure difference is shown at $\frac{1}{8}$ msec intervals (every $\pi/4$ radians of phase) for one cycle. The traveling wave nature of the excitation is apparent, with the speed of propagation along the membrane being greater at the basal end and becoming slower as the apex (helicotrema) is approached.

From the pressure and velocity solutions, an equivalent four-pole network can be deduced for an incremental length of the cochlea. Voltage can be taken analogous to sound pressure and current analogous

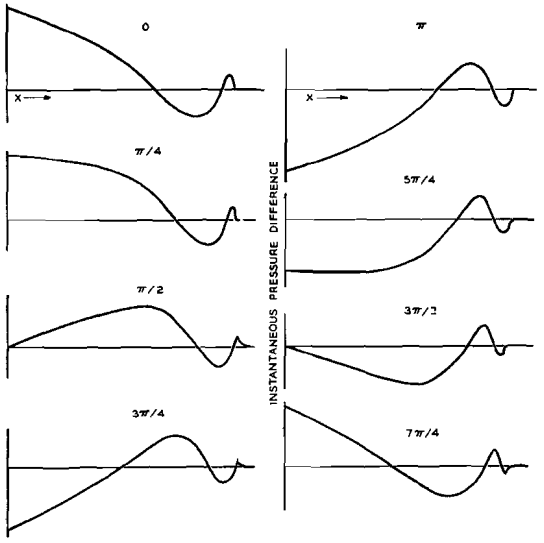

Fig. 4.35. Instantaneous pressure difference across the cochlear partition at successive phases in one period of a 1000 cps excitation. (After PETERSON and BOGERT)

to volume velocity. Such a network section is shown in Fig. 4.36 (BOGERT). Here L_1 represents the mass of the fluid in an incremental length of the scalas; C_1 the compressibility of the fluid; and L_2, R_1, C_2, C_3, and C_4 represent the mechanical constants of the membrane. The voltage $P(x, \omega)$ represents the pressure difference across the membrane as a function of distance and frequency, and the voltage $Y(x, \omega)$ represents the membrane displacement.

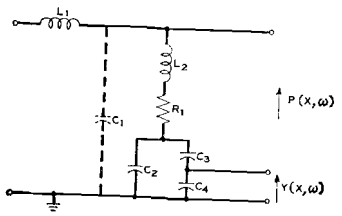

Fig. 4.36. Electrical network section for representing an incremental length of the cochlea. (After BOGERT)

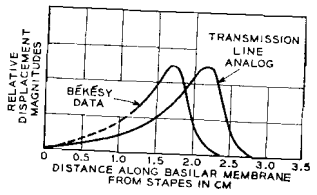

Fig. 4.37. Comparison of the displacement response of the transmission line analog of the cochlea to physiological data for the ear. (After BOGERT)

A set of 175 such sections has been used to produce a transmission line analog of the cochlea (BOGERT). The displacement responses exhibited by the line compare well in shape with those measured by BÉKÉSY on real cochleas. An illustrative response is shown in Fig. 4.37. Some differences are found in the positions of peak response and in the lowest frequencies which exhibit resonance phenomena. Probable origins of the differences are the uncertainties connected with the spatial variation of the measured mechanical constants of the membrane and the neglect of mutual coupling among membrane elements. Despite the uncertainties in the distributed parameters, the transmission line analog provides a graphic demonstration of the traveling-wave nature of the basilar membrane motion.

4.3. Illustrative Relations between Subjective and Physiological Behavior

The ear models discussed above describe only the mechanical functions of the peripheral ear. Any comprehensive hypothesis about auditory perception must provide for the transduction of mechanical displacement into neural activity. As indicated earlier, the details of this process are not well understood. The assumptions that presently can be made are of a gross and simplified nature. Three such assumptions are useful,

however, in attempting to relate physiological and subjective behavior. Although oversimplifications, they do not seem to violate known physiological facts.

The first is that sufficient local deformation of the basilar membrane elicits neural activity in the terminations of the auditory nerve. A single neuron is presumably a binary (fired or unfired) device. The number of neurons activated depends in a monotonic fashion upon the amplitude of membrane displacement[1]. Such neural activity may exist in the form of volleys triggered synchronously with the stimulus, or in the form of a signalling of place localization of displacement. Implicit is the notion that the displacement—or perhaps spatial derivatives of displacement—must exceed a certain threshold before nerve firings take place.

Second, neural firings occur on only one "polarity" of the membrane displacement, or of its spatial derivative. In other words, some process like half-wave rectification operates on the mechanical response. Third, the membrane point displacing with the greatest amplitude originates the predominant neutral activity. This activity may operate to suppress or inhibit activity arising from neighboring points.

These assumptions, along with the results from the models, have in a number of instances been helpful in interpreting auditory subjective behavior. Without going into any case in depth, several applications can be outlined.

4.31. Pitch Perception

Pitch is that subjective attribute which admits of a rank ordering on a scale ranging from low to high. As such, it correlates strongly with objective measures of frequency. One important facet of auditory perception is the ability to ascribe a pitch to sounds which exhibit periodic characteristics.

Consider first the pitch of pure (sinusoidal) tones. For such stimuli the basilar membrane displacements are, of course, sinusoidal. The frequency responses given previously in Fig. 4.27a indicate the relative amplitudes of displacement versus frequency for different membrane points. At any given frequency, one point on the membrane responds with greater amplitude than all others. In accordance with the previous assumptions, the most numerous neural volleys are elicited at this maximum point. For frequencies sufficiently low (less than about 1000 cps), the volleys are triggered once per cycle and at some fixed epoch on the displacement waveform. Subsequent processing by higher

[1] Psychological and physiological evidence suggests that the intensity of the neural activity is a power-law function of the mechanical displacement. A single neuron is also refractory for a given period after firing. A limit exists, therefore, upon the rate at which it can fire.

centers presumably appreciates the periodicity of the stimulus-locked volleys.

For frequencies greater than about 1000 to 2000 cps, electro-physiological evidence suggests that synchrony of neural firings is not maintained (GALAMBOS). In such cases, pitch apparently is perceived through a signalling of the place of greatest membrane displacement. The poorer frequency resolution of points lying in the basal part of the basilar membrane probably also contributes to the psychoacoustic fact that pitch discrimination is less acute at higher frequencies.

Suppose the periodic sound stimulus is not a simple sinusoidal tone but is more complex, say repeated sharp pulses. What pitch is heard? For purpose of illustration, imagine the stimulus to be the alternately positive and negative impulses used to illustrate the digital simulation in Fig. 4.33. Such a pulse train has a spectrum which is odd-harmonic. If the pulses occur slowly enough, the membrane displacement at all points will resolve each pulse in time. That is, the membrane will have time to execute a complete, damped impulse response at all places for each pulse, whether positive or negative. Such a situation is depicted by the analog membrane responses shown in the left column of Fig. 4.38. The fundamental frequency of excitation is 25 cps (50 pps). The waveforms were measured from analog networks such as illustrated in Fig. 4.28.

For this low pulse rate condition, one might imagine that neural firings synchronous with each pulse—regardless of polarity—would be triggered at all points along the membrane. The perceived pitch might

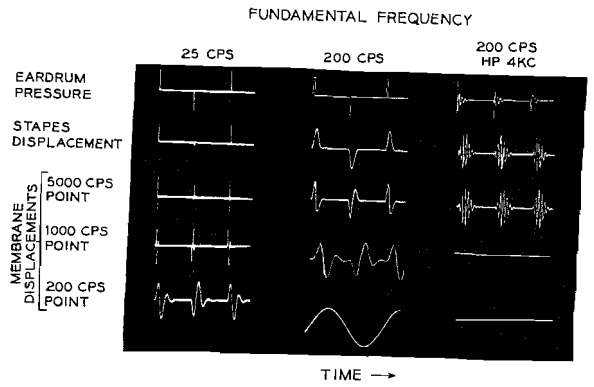

Fig. 4.38. Membrane displacement responses for filtered and unfiltered periodic pulses. The stimulus pulses are alternately positive and negative. The membrane displacements are simulated by the electrical networks shown in Fig. 4.28. To display the waveforms more effectively, the traces are adjusted for equal peak-to-peak amplitudes. Relative amplitudes are therefore not preserved

then be expected to be equal to the pulse rate. Measurements show this to be the case (FLANAGAN and GUTTMAN). Furthermore, the model indicates that a pulse signal of this low rate causes the greatest displacements near the middle portion of the membrane, that is, in the vicinity of the place maximally responsive to about 1500 cps.

If, on the other hand, the fundamental frequency of excitation is made sufficiently high, say 200 cps or greater, the fundamental component will be resolved (in frequency) at the most apically-responding point. This situation is illustrated for a 200 cps fundamental by the traces in the second column of Fig. 4.38. The 200 cps place on the membrane displaces with a nearly pure sinusoidal motion, while the more basal points continue to resolve each pulse in time. At the apical end, therefore, neural volleys might be expected to be triggered synchronously at the fundamental frequency, while toward the basal end the displacements favor firings at the pulse rate, that is, twice per fundamental period. Psychoacoustic measurements indicate that the apical, fundamental-correlated displacements are subjectively more significant than the basal, pulse-rate displacements. The fundamental-rate volleys generally predominate in the percept, and the pitch is heard as 200 sec^{-1}. At some frequency, then, the pitch assignment switches from pulse rate to fundamental.

The pulse pattern illustrating the computer simulation in Fig. 4.33 is the same positive-negative pulse alternation under discussion, but it is produced at a fundamental frequency of 100 cps. This frequency is immediately in the transition range between the fundamental and pulse-rate pitch modes. One notices in Fig. 4.33 that the ear is beginning to resolve the fundamental component in relatively low amplitude at the apical end of the membrane, while the pulse rate is evident in the basal displacements. One might suppose for this condition that the pulse rate *and* fundamental cues are strongly competing, and that the pitch percept is ambiguous. Subjective measurements bear this out.

Another effect becomes pronounced in and near the pitch-transition region corresponding to the conditions of Fig. 4.33. A fine structure in the perception of pulse pitch becomes more evident. The membrane region where displacement amplitude is greatest is in the place-frequency range 600 to 1500 cps. In this region the displacement response to a pulse has a period which is an appreciable fraction of the pulse repetition period. That is, the half-period time of the pulse response is a significant percentage of the pulse period. Assume as before that neural firings occur only on positive deflections of the membrane. The intervals between firings on fibers originating from a given place in this region should, therefore, be alternately lengthened and shortened. The change in interval (from strict periodicity) is by an amount equal to the

half-period of the pulse response at that place. One might expect, therefore, a bimodality in the pitch percept. If f_d is the place-frequency of dominant membrane motion and r the signal pulse rate, the perceived pitch f_p should correspond to

$$f_p = \left[\frac{1}{r} \pm \frac{1}{2f_d}\right]^{-1}.$$

This bimodality in the pitch percept is in fact found (ROSENBERG; RITSMA).

If the 200 cps stimulus in the middle column of Fig. 4.38 is high-pass filtered at a sufficiently high frequency, only the basal displacements remain effective in producing the pitch percept. For example, the membrane displacements for a high-pass filtering at 4000 cps are shown in the third column of Fig. 4.38. If the present arguments continue to hold, such a filtering should change the percept from the fundamental mode back to the pulse-rate mode. The reason, of course, is that the time resolution of the basal end separates each pulse, whether positive or negative. This hypothesis is in fact sustained in psychoacoustic measurements (GUTTMAN and FLANAGAN, 1964).

A somewhat more subtle effect is obtained if the high-pass filtering is made at a fairly small harmonic number, for example, at the second harmonic, so as to remove only the fundamental component. Under certain of these conditions, the membrane may generate displacements which favor a difference-frequency response. For a stimulus with odd and even components, the pitch percept can be the fundamental, even though the fundamental is not present in the stimulus.

4.32. Binaural Lateralization

Another aspect of perception is binaural lateralization. This is the subjective ability to locate a sound image at a particular point inside the head when listening over earphones. If identical clicks (impulses of sound pressure) are produced simultaneously at the two ears, a normal listener hears the sound image to be located exactly in the center of his head. If the click at one ear is produced a little earlier or with slightly greater intensity than the other, the sound image shifts toward that ear. The shift continues with increasing interaural time or intensity difference until the image moves completely to one side and eventually breaks apart. One then begins to hear individual clicks located at the ears.

Naively we suppose the subjective position of the image to be determined by some sort of computation of coincidence between neural volleys. The volleys originate at the periphery and travel to higher centers via synaptic pathways. The volley initiated earliest progresses

to a point in the neural net where a coincidence occurs with the later volley. A subjective image appropriately off-center is produced. To the extent that intensity differences can shift the image position, intensity must be coded – at least partially – in terms of volley timing. As has been the case in pitch perception, there are several research areas in binaural phenomena where the computational model described in Section 4.2 has been helpful in quantifying physiological response and relating it to subjective behavior. One such area concerns the effects of phase and masking upon the binaural lateralization of clicks.

If a pulse of pressure rarefaction is produced at the eardrum, the drum is initially drawn outward. The stapes is also initially drawn outward, and the membrane is initially drawn upward. The stapes and membrane displacements (as described by the model) in response to a rarefaction pulse of 100 μsec duration are shown by the waveforms at the right of Fig. 4.39. The pulse responses of three different membrane points are shown, namely, the points maximally responsive to 2400 cps, 1200 cps, and 600 cps, respectively. The stapes displacement is a slightly integrated version of the input. The membrane responses reflect the vibratory behavior of the particular points as well as the traveling-wave transit delay to the points.

According to the model, broadband pulses produce the greatest displacements near the middle of the membrane, roughly in the region

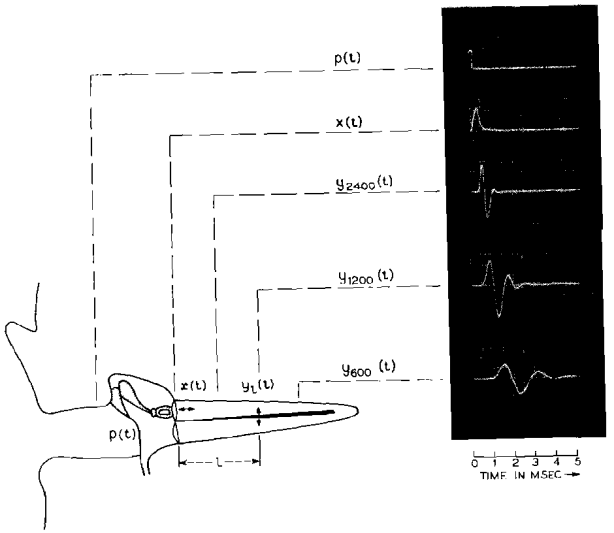

Fig. 4.39. Basilar membrane responses at the 2400, 1200 and 600 cps points to a pressure-rarefaction pulse of 100 μsec duration. The responses are measured on the electrical analog circuit of Fig. 4.28. Relative amplitudes are preserved

maximally responsive to about 1 500 cps. The magnitude of displacement is less at places either more toward the base or more toward the apex. It has been hypothesized that the most significant neural activity is generated at the membrane point displacing with the greatest amplitude. Further, electro-physiological data suggest that neural firings occur at some threshold only on unipolar motions of the basilar membrane. (For the outer hair cells, these are motions which drive the basilar membrane toward the tectorial membrane.) The oscillatory behavior of the pulse response suggests, too, that multiple or secondary neural firings might be elicited by single stimulus pulses.

If pulses are supplied to both ears, a centered sound image is heard if the significant neural activity is elicited simultaneously. Suppose that the input pulses are identical rarefaction pulses. The maximum displacements occur near the middle of the membrane. For simplicity imagine that the neural firings are triggered somewhere near the positive crests of the displacement waves. For this cophasic condition, a centered image is heard if the input pulses are produced simultaneously, or if the interaural time is zero. Suppose now that the pulse to one of the ears is reversed in phase to a pressure condensation. The membrane responses for this ear also change sign and are the negatives of those shown in Fig. 4.39. Their first positive crests now occur later by about one-half cycle of the displacement at each point. At the middle of the membrane this half-cycle amounts to about 300 to 400 μsec. To produce a centered image for the antiphasic condition, then, one would expect that the condensation pulse would have to be advanced in time by this amount.

The membrane point which displaces with the greatest coherent amplitude can be manipulated by adding masking noise of appropriate frequency content. That is, the place which normally responds with greatest amplitude can be obscured by noise, and the significant displacement caused to occur at a less sensitive place. For example, suppose that the basal end of the membrane in one ear is masked by high-pass noise, and the apical end of the membrane in the other ear is masked by low-pass noise. If the listener is required to adjust stimulus pulses to produce a centered image, the fusion must be made from apical-end information in one ear and basal-end in the other. The resulting interaural time would then reflect both the oscillatory characteristics of the specific membrane points and the traveling-wave delay between them.

Experiments show these time dependencies to be manifest in subjective behavior (FLANAGAN, DAVID, and WATSON). The test procedure to measure them is shown in Fig. 4.40. Identical pulse generators produce 100 μsec pulses at a rate of 10 per second. Pulse amplitude is set to produce a 40 db sensation level. The subject, seated in a sound-

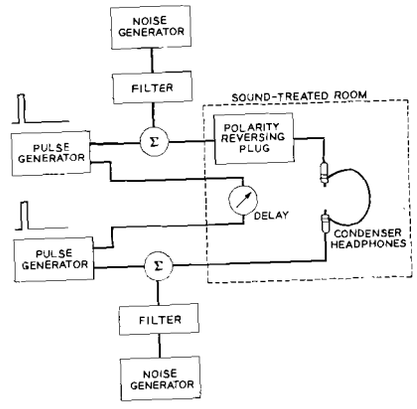

Fig. 4.40. Experimental arrangement for measuring the interaural times that produce centered sound images. (After FLANAGAN, DAVID and WATSON)

treated room, listens to the pulses over condenser earphones. (Condenser phones are used because of the importance of good acoustic reproduction of the pulses.) He has a switch available to reverse the polarity of the pulses delivered to the right ear so that it can be made a condensation instead of the normal rarefaction. The subject also has a delay control which varies the relative times of occurrence of the two pulses over a range of ± 5 msec. Two uncorrelated noise generators supply masking noise via variable filters. (A separate experiment was conducted to determine the filtered noise levels necessary to mask prescribed spectral portions of the pulse stimuli.)

For a given masking and pulse polarity condition, the subject is required to adjust the delay to produce a centered sound image in his head. Multiple images are frequently found, with the more subtle, secondary images apparently being elicited on secondary bounces of the membrane.

Fig. 4.41 shows the results for principal-image fusions under a variety of masking conditions. Fig. 4.41 a gives results for unmasked and symmetrically-masked conditions, and Fig. 4.41 b gives the results for asymmetrical masking. The data are for four subjects, and each point is the median of approximately 15 principal-image responses. Each two sets of points is bracketed along the abscissa. The set labelled *C* is the cophasic response and that labelled *A* is the antiphasic. The cophasic conditions are rarefaction pulses in both ears. The antiphasic conditions are rarefaction in the left ear and condensation in the right ear.

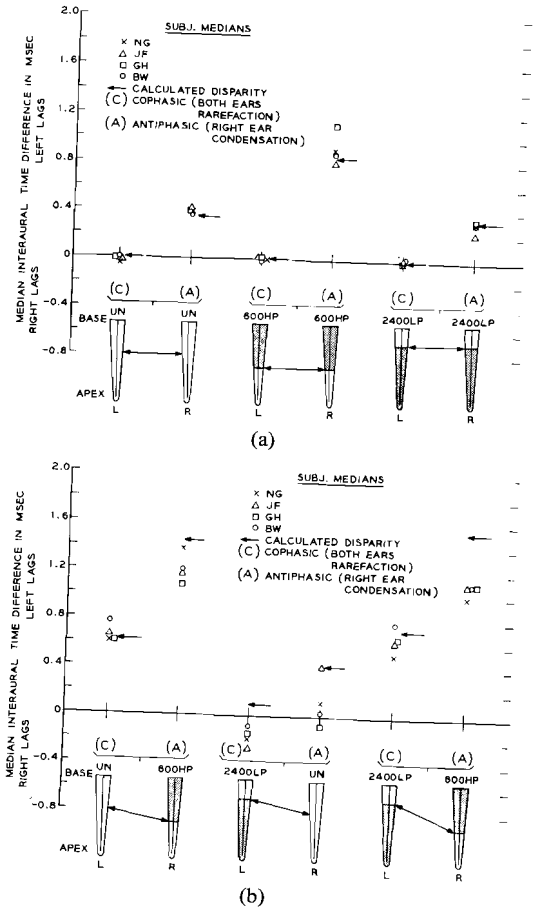

Fig. 4.41 a and b. Experimentally measured interaural times for lateralizing cophasic and antiphasic clicks. Several conditions of masking are shown. (a) Unmasked and symmetrically masked conditions. (b) Asymmetrically masked conditions. The arrows indicate the interaural times predicted from the basilar membrane model

Each bracket corresponds to the masking conditions represented by the schematic cochleas drawn below the brackets. The labelling at the top of each cochlea gives the masking condition for that ear. For example, the UN means unmasked. The dark shading on the cochleas indicates the membrane regions obscured by masking noise. The double arrow between each pair of cochleas indicates approximately the points of maximum, unmasked displacement. For example, in the first case of Fig. 4.41 a, which is the unmasked case, the maximum displacements occur near the middles of the two membranes.

The single arrows in the vicinity of the plotted responses are estimates of the interaural times calculated from the basilar membrane model. The estimates are made by assuming the neural firings to be produced at the positive crest of the displacement at the most significant place. The arrows therefore represent the time differences between the first positive crests at the places indicated in the cochlear diagrams. As such, they include the transit time to the particular place, plus the initial quarter-cycle duration of the pulse response.

The actual threshold for neural firing is of course not known, and is very likely to be dependent upon place. In the symmetrically-masked conditions, an actual knowledge of the threshold is not of much consequence since the threshold epoch, whether it is at the crest or down from the crest, should be about the same in the two ears. For these cases, therefore, it is the half-cycle time of the displacement wave that is important. Fig. 4.41 a shows that the measured responses do, in fact, agree relatively well with this simple estimate of the interaural time. All of the principal cophasic fusions are made for essentially zero time, and the antiphasic lateralizations reflect the half-cycle disparity of the appropriate places, with the condensation pulse always leading.

The agreement is not as good for the asymmetrically-masked cases shown in Fig. 4.41 b. Signal loudnesses are different in the two ears, and the neural thresholds probably vary with place. The times of the initial positive crests would not be expected to give very realistic estimates of the interaural times. It becomes much more important to have a knowledge of the actual threshold levels and the relative amplitudes of the displacements. Even so, it is interesting to note to what extent the simple positive-crest estimates follow the data.

In the first condition, the left ear is unmasked and the right ear has masking noise high-pass filtered at 600 cps (600 HP). The cophasic interaural time is predicted to be on the order of 600 μsec, and the measurements do give essentially this figure. The antiphasic condition is expected to be on the order or 1450 μsec, but the measured median response is a little less, about 1200 μsec.

The next case has the left ear masked with noise low-pass filtered at 2400 cps (2400 LP), and the right ear is unmasked. The cophasic condition is expected to yield an interaural time of slightly less than 100 μsec, with the left ear lagging, but the experimental measurements actually give a right ear lag of about 150 μsec. The relatively wide spread of the subject medians in the asymmetrical cases, and in particular for the cases involving 2400 LP, show that these lateralizations are considerably more difficult and more variable than the symmetrical cases. The antiphasic response for this same condition is estimated to give an interaural time on the order of 400 μsec, but again the responses are variable with the

median falling at about 100 μsec. One subject's median actually falls on the right-lag side of the axis.

The final condition has 2400 LP in the left ear and 600 HP in the right ear. The cophasic fusion is expected to be in the neighborhood of 700 μsec, and the measured response is found about here. The antiphasic condition should yield an interaural time on the order of 1550 μsec, but the measurements produce a median slightly greater than 1100 μsec.

Clearly, the simple assumption of neural firing at the positive crests (or some other fixed epoch) of the displacement is not adequate to specify all of the interaural times. The real thresholds for firing are likely to vary considerably with place. In fact, by taking data such as these, plus the displacement waves from the model, the possibility exists for working backwards to deduce information about neural threshold epochs. More broadly than this, however, the present results suggest strong ties between subjective response and the detailed motion of the basilar membrane.

4.33. Threshold Sensitivity

The combined response curves in Fig. 4.27a indicate that the ear is mechanically more sensitive to certain frequencies than to others. A similar frequency dependence exists subjectively. To what extent are the variations in the threshold of audibility accounted for simply by the mechanical sensitivity of the ear?

The envelope of the peak displacement responses in Fig. 4.27a can be compared with the subjectively determined minimum audible pressure for pure (sine) tones. Fig. 4.42 shows this comparison. The agreement

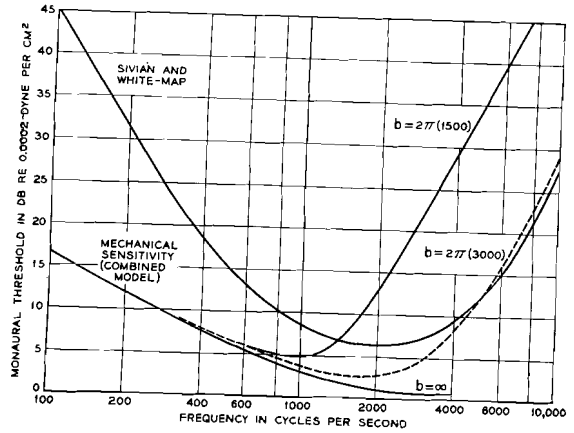

Fig. 4.42. Relation between the mechanical sensitivity of the ear and the monaural minimum audible pressure threshold for pure tones

is generally poor, although the gross trends are similar. One curve in the figure is based on the 1500 cps critical frequency for the middle ear. The earlier discussion has pointed up the uncertainty and variability of this figure. If a critical frequency of 3000 cps is taken for the middle ear, the fit to the threshold curve at high frequencies is more respectable[1]. The match at low frequencies, however, is not improved, but this is of less concern for a different reason.

At the low frequencies, the disparity between the mechanical and subjective sensitivity may be partially a neural effect. According to the earlier assumptions, the number of neurons activated bears some monotonic relation to amplitude of membrane displacement. Perception of loudness is thought to involve possibly temporal and spatial integrations of neural activity. If a constant integrated neural activity were equivalent to constant loudness, the difference between mechanical and subjective sensitivities might be owing to a sparser neural density in the apical (low-frequency) end of the cochlea. There is physiological evidence to this effect.

In histological studies, GUILD *et al.* counted the number of ganglion cells per mm length of the organ of Corti. Their results for normal ears are summarized in Fig. 4.43. These data show a slight decrease in the number of cells at the basal end, and a substantial decrease in the density as the apex is approached. The innervation over the middle of the membrane is roughly constant.

One can pose similar questions about threshold sensitivity to short pulses or clicks of sound. For pulses of sufficiently low repetition rate, the maximal displacement of the membrane—as stated before—is near

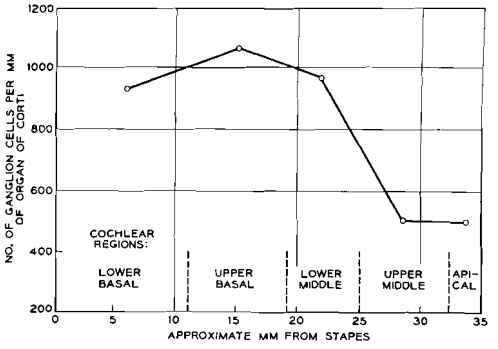

Fig. 4.43. Average number of ganglion cells per mm length of organ of Corti. (After GUILD *et al.*)

[1] Note, too, that the membrane velocity response $\dot{y}(t)$ provides a better fit to the tone threshold than does the displacement, $y(t)$. $\dot{y}(t)$ includes an additional +6 db/oct. component.

the middle. According to the model, the place of maximum displacement remains near the middle for pulse rates in excess of several hundred per second. In its central region, the resonance properties of the membrane permit temporal resolution of individual exciting pulses for rates upwards of 1000 sec^{-1}. If the predominant displacement takes place in one vicinity for a large range of pulse rates, polarity patterns and pulse durations, how might the subjective threshold vary with these factors, and how might it be correlated with the membrane motion? At least one examination of this question has been made (FLANAGAN, 1961a). The results can be briefly indicated.

Binaural thresholds of audibility for a variety of periodic pulse trains with various polarity patterns, pulse rates and durations are shown in Fig. 4.44. The data show that the thresholds are relatively independent of polarity pattern. For pulse rates less than 100 pps, the thresholds are relatively independent of rate, and are dependent only upon duration. Above 100 pps, the thresholds diminish with increasing pulse rate. The amplitude of membrane displacement would be expected to be a function of pulse duration and to produce a lower threshold for the longer pulses, which is the case. For rates greater than 100 sec^{-1}, however, some other nonmechanical effect apparently is of importance. The way in which audible pulse amplitude diminishes suggests a temporal integration with a time constant of the order of 10 msec.

Using the earlier assumptions about conversion of mechanical to neural activity, one might ask "what processing of the membrane displacement at the point of greatest amplitude would reflect the constant loudness percept at threshold." One answer is suggested by the operations illustrated in Fig. 4.45. The first two blocks represent middle ear transmission [as specified in Eq. (4.3)] and basilar membrane displacement [vicinity of the 1000 cps point, as specified in Eq. (4.1)]. The

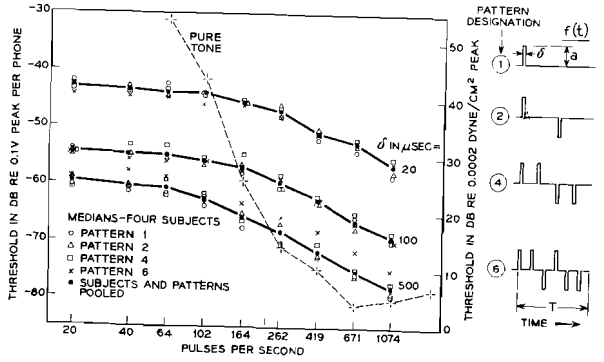

Fig. 4.44. Binaural thresholds of audibility for periodic pulses. (After FLANAGAN, 1961a)

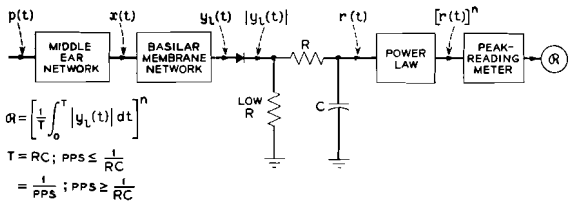

Fig. 4.45. Model of the threshold of audibility for the pulse data shown in Fig. 4.44

diode represents the half-wave rectification associated with neural firings on unipolar motions of the membrane. The *RC* integrator has a 10 msec time constant, as suggested by the threshold data. The power-law element (exponent = 0.6) represents the power-law relation found in loudness estimation[1]. A meter indicates the peak value of the output of the power-law device. When the stimulus conditions represented by points on the threshold curves in Fig. 4.44 are applied to this circuit, the output meter reads the same value, namely, threshold.

One can also notice how this simple process might be expected to operate for sine wave inputs. Because the integration time is 10 msec, frequencies greater than about 100 cps produce meter readings proportional to the average value of the half-wave rectified sinusoid. In other words, the meter reading is proportional to the amplitude of the sine wave into the rectifier. Two alterations in the network circuitry are then necessary. First, the basilar membrane network appropriate to the point maximally responsive to the sine frequency must be used. This may be selected from an ensemble of networks. And second, to take account of the sparser apical innervation, the signal from the rectifier must be attenuated for the low-frequency networks in accordance with the difference between the mechanical and subjective sensitivity curves in Fig. 4.42. The power-law device is still included to simulate the growth of loudness with sound level.

4.34. Auditory Processing of Complex Signals

The preceding discussions suggest that the extent to which subjective behavior can be correlated with (and even predicted by) the physiological operation of the ear is substantial. Recent electrophysiological data link neural activity closely with the detailed mechanical motion of the basilar membrane. Subjective measurements, such as described in the foregoing

[1] The power-law device is not necessary for simple threshold indications of "audible-inaudible". It is necessary, however, to represent the growth of loudness with sound level, and to provide indications of subjective loudness above threshold.

sections, lend further support to the link. Psychological and physiological experimentation continue to serve jointly in expanding knowledge about the processes involved in converting the mechanical motions of the inner ear into intelligence-preserving neural activity.

The physiological-psychoacoustic correlations which have been put forward here have involved only the simplest of signals – generally, signals that are temporally punctuate or spectrally discrete, or both. Furthermore, the correlations have considered only gross and salient features of these signals, such as periodicity or time of occurrence. The primary aim has been to outline the peripheral mechanism of the ear and to connect it with several phenomena in perception. Little has been said about classical psychoacoustics or about speech perception. As the stimuli are made increasingly complex – in the ultimate, speech signals – it seems clear that more elaborate processing is called into play in perception. Much of the additional processing probably occurs centrally in the nervous system. For such perception, the correlations that presently can be made between the physiological and perceptual domains are relatively rudimentary. As research goes forward, however, these links will be strengthened.

The literature on hearing contains a large corpus of data on subjective response to speech and speech-like stimuli. There are, for example, determinations of the ear's ability to discriminate features such as vowel pitch, loudness, formant frequency, spectral irregularity and the like. Such data are particularly important in establishing criteria for the design of speech transmission systems and in estimating the channel capacity necessary to transmit speech data. Instead of appearing in this chapter, comments on these researches have been reserved for a later, more applied discussion where they have more direct application to transmission systems.

V. Techniques for Speech Analysis

The earlier discussion suggested that the encoding of speech information might be considered at several stages in the communication chain. On the transmitter side, the configuration and excitation of the vocal tract constitute one description. In the transmission channel, the transduced acoustic waveform is a signal representation commonly encountered. At the receiver, the mechanical motion of the basilar membrane is still another portrayal of the information. Some of these descriptions exhibit properties which might be exploited in communication.

Efforts in speech analysis and synthesis frequently aim at the efficient encoding and transmission of speech information[1]. Here the goal is the transmission of speech information over the smallest channel capacity adequate to satisfy specified perceptual criteria. Acoustical and physiological analyses of the vocal mechanism suggest certain possibilities for efficient description of the signal. Psychological and physiological experiments in hearing also outline certain bounds on perception. Although such analyses may not necessarily lead to totally optimum methods for encoding and transmission, they do bring to focus important physical constraints. Transmission economies beyond this level generally must be sought in linguistic and semantic dependencies.

The discussions in Chapters II and III set forth certain fundamental relations for the vocal mechanism. Most of the analyses presumed detailed physical knowledge of the tract. In actual communication practice, however, one generally has knowledge only of some transduced version of the acoustic signal. (That is, the speaker does not submit to measurements on his vocal tract.) The acoustic and articulatory parameters of the preceding chapters must therefore be determined from the speech signal if they are to be exploited.

This chapter proposes to discuss certain speech analysis techniques which have been found useful for deriving so-called "information-bearing elements" of speech. Subsequent chapters will consider synthesis of speech from these low information-rate parameters, perceptual criteria appropriate to the processing of such parameters, and application of analysis, synthesis and perceptual results in complete transmission systems.

5.1. Spectral Analysis of Speech

Frequency-domain representation of speech information appears advantageous from two standpoints. First, acoustic analysis of the vocal mechanism shows that the normal mode or natural frequency concept permits concise description of speech sounds. Second, clear evidence exists that the ear makes a crude frequency analysis at an early stage in its processing. Presumably, then, features salient in frequency analysis are important in production and perception, and consequently hold promise for efficient coding. Experience supports this notion.

Further, the vocal mechanism is a quasi-stationary source of sound. Its excitation and normal modes change with time. Any spectral measure applicable to the speech signal should therefore reflect temporal features of perceptual significance as well as spectral features. Something other then a conventional frequency transform is indicated.

[1] Other motivating objectives are: basic understanding of speech communication, voice control of machines, and voice response from computers.