

## A SPACE-TIME PATTERN THEORY OF HEARING

BY HARVEY FLETCHER

*Bell Telephone Laboratories*

In constructing any theory of hearing, six classes of quantitative information must be considered, namely, (1) the limits of the pitch and the intensity that the hearing mechanism is capable of handling, (2) the minimum changes in pitch and intensity that are perceptible, (3) subjective tones, (4) the masking effects of one sound upon another, (5) the relation of loudness to sensation level, and (6) binaural effects. Besides these effects there are other auditory phenomena which cannot be expressed in as quantitative a manner but they are just as important. The principal ones are (1) the ability of the ear to analyze a complex tone into its component tones, (2) the effect upon the pitch, the loudness, and the quality of a musical tone when certain groups of its components are eliminated, (3) the relation of brightness and volume (as used by psychologists) to frequency and intensity of the stimulating tone, and (4) the effect of shifting the phrases of the components of a musical tone. It is assumed that the reader is familiar with these various effects. They will be referred to at the appropriate time as the theory of hearing is developed.

Two general types of hearing theories have been put forth from time to time to explain these effects. One might be called a space pattern theory and the other a time pattern theory. In the first theory, it is assumed that the time pattern of the wave motion in the air is transferred into a space pattern in the inner ear so that the nerve impulses reaching the brain give us information concerning the time pattern of the wave motion by means of the location of the nerves which are stimulated. In the second theory, it is assumed that the time sequences are transmitted directly to the brain. It is the opinion of the author that both of these effects are operating in aiding one to interpret the sounds which one hears. The term "A Space-Time Pattern Theory of Hearing" therefore best expresses this conception.

When a sound wave is created in the air, the air particles oscillate back and forth and cause small pressure changes which occur in a certain time sequence depending upon the kind of sound created.

If a small<sup>1</sup> pressure gauge which follows instantly the pressure varia-

<sup>1</sup> This size must not be too small or the Brownian Movements would be superimposed upon the wave motion.

tions of the air, were placed at a certain point in the path of the wave, its fluctuations for the three classes of sound indicated, would be as shown in Fig. 1. Each curve is a pressure time pattern of the sound it represents. This pattern is transmitted through the air and communicated to the drum of the ear. It is then transferred through

### TIME PATTERNS

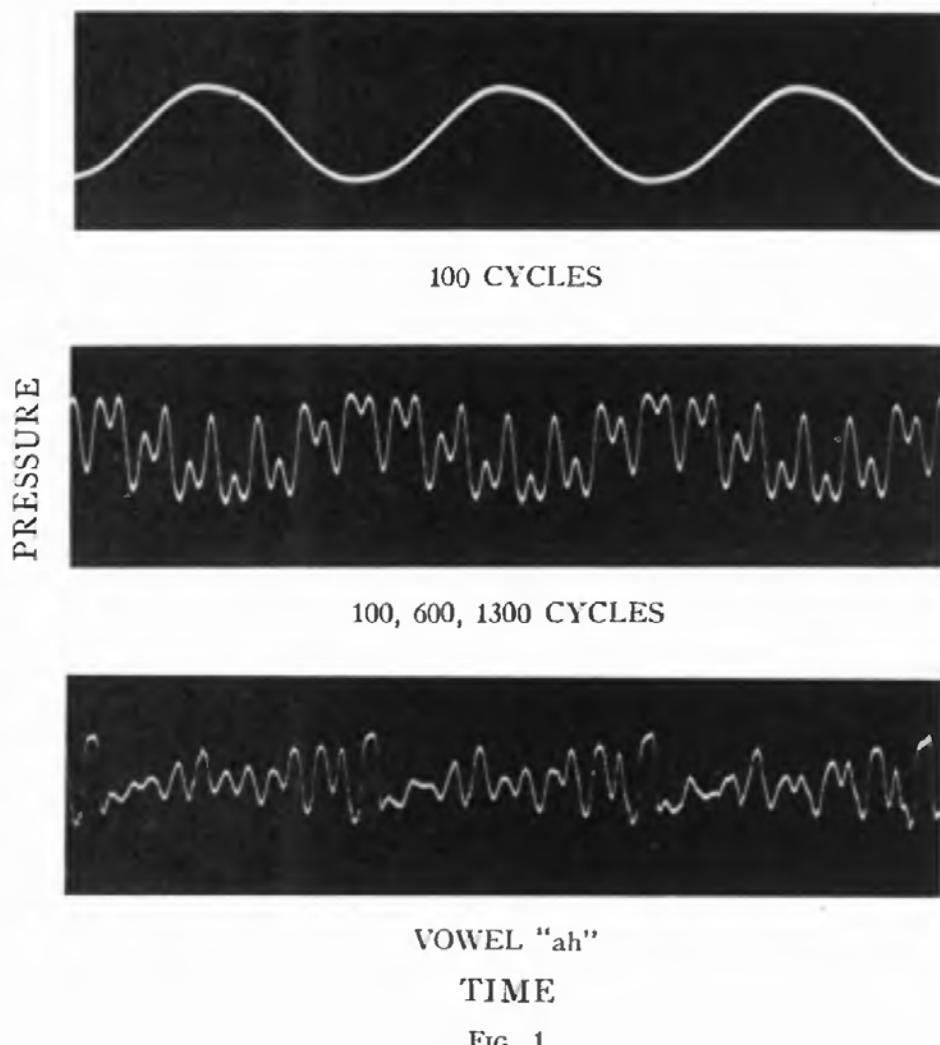


FIG. 1

the ossicles of the middle ear to the liquid of the cochlea. In this process of transmission, the pressure changes are usually considered to be magnified about 60 times. This is due to the lever action of the ossicles and also because the total force<sup>2</sup> exerted upon the ear drum must be distributed over the much smaller area of the oval window. If we knew

<sup>2</sup> This is only true provided that the ear drum does not break up into segments as it vibrates.

all of the mechanical constants involved we could predict the other changes in the pattern. For low intensities and for sounds containing components having vibration frequencies in the middle range, the time pattern probably undergoes only slight modifications. Components having frequencies above or below this range are probably transmitted less efficiently. In any case there will exist a definite time-pressure pattern in the liquid of the cochlea near the oval window which is similar to but not exactly of the same form as that existing in the air.

#### NON-LINEAR CHARACTERISTICS OF THE MIDDLE EAR

If the intensity of the sound is increased without otherwise changing the pattern in the air, there is a change in the pattern transmitted to the cochlea not only in magnitude but also in form. For high sound intensities the elastic members in the middle ear will exceed the amplitude limits within which the strain is proportional to the stress. Also, the limit for extension is different from that for compression. Under such circumstances it is well known that frequencies other than those in the original pattern are introduced. For example, for such high intensities a pure tone of frequency  $f$  will be transmitted to the cochlea as a tone having components of frequency  $f$ ,  $2f$ ,  $3f$ ,  $4f$ , etc. These components which are added to the original tone in the process of transmission account for the subjective tones. Their magnitude can be measured experimentally as follows. Introduce the pure tone into the ear at a known intensity level. At the same time introduce another external tone which differs in frequency from the harmonic being measured by two or three cycles. Beats will then be produced between the exploring tone and the subjective harmonic. If the intensity of the exploring tone is adjusted until the most distinct beats are obtained, then its intensity is a measure of the magnitude of the subjective harmonic. In Fig. 2 are shown the magnitudes of such harmonics for various intensities of the stimulating sound. These curves were constructed from some experimental measurements made by Mr. F. H. Graham. The abscissae give the number of the harmonic<sup>3</sup> and the ordinates the intensity level. For example, a tone having a fundamental<sup>3</sup> at zero level (1 microwatt per square centimeter) will produce subjective harmonics having levels of  $-1.4$ ,  $-2.6$ ,  $-3.7$ , and  $-4.8$  bels for the second, third, fourth, and fifth harmonics, respectively. It is difficult to determine the exact intensity of the subjective tones, yet within the

<sup>3</sup> The fundamental is called the first harmonic.

observational error and individual variation, it is found that this set of curves fits the experimental facts for tones of all frequencies. This is what one might expect, since the overloading of the middle ear depends upon the physical intensity of the impressed sound rather than upon its sensation level.

Further it might be expected that this overloading would be dependent upon the amplitude of vibration of the ear drum and the chain

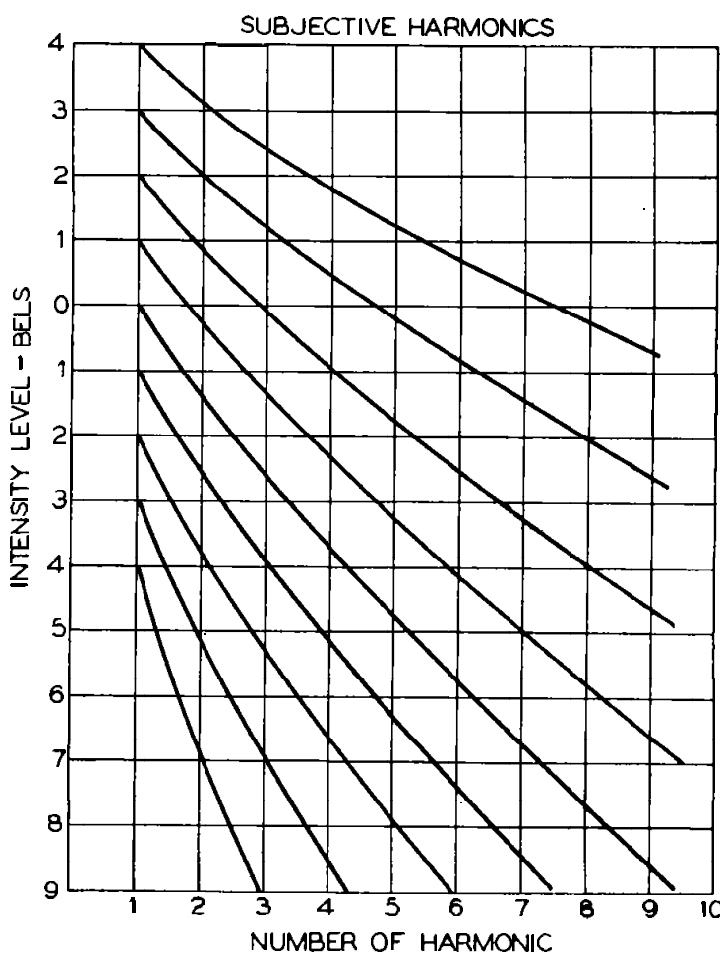


FIG. 2

of bones attached to it. To be in agreement with the above experimental facts it is evident that the impedance at the ear drum must be largely in the nature of a stiffness rather than a resistance or a mass. Recent data by Tröger<sup>4</sup> show that the impedance of the ear is of this character except for high frequencies. It is important to remember that the curves in Fig. 2 are for intensity levels rather than for sensa-

<sup>4</sup> The Reception of Sound by the Outer Ear, Phys. Zeit., Vol. 31, no. 1, Jan. 1930, p. 26-47.

tion levels. Expressed in terms of sensation levels the magnitudes of the subjective harmonics for the tone given above, when its pitch is  $-4$  octaves, would be  $4.4, 4.6, 4.4, 3.8$ , and  $3.0$  bels above the threshold. It is seen that the first harmonic has a higher sensation level than the fundamental. For still higher intensity levels several of the harmonics may have higher sensation levels than the fundamental.

In Fig. 3 are shown the observational points for four different levels of a tone having a frequency of 50 cycles per second. As you will notice the first six overtones for the higher levels have a sensation level higher than the fundamental. It must be remembered that the data are limited and tests upon any one ear give results which differ appreciably

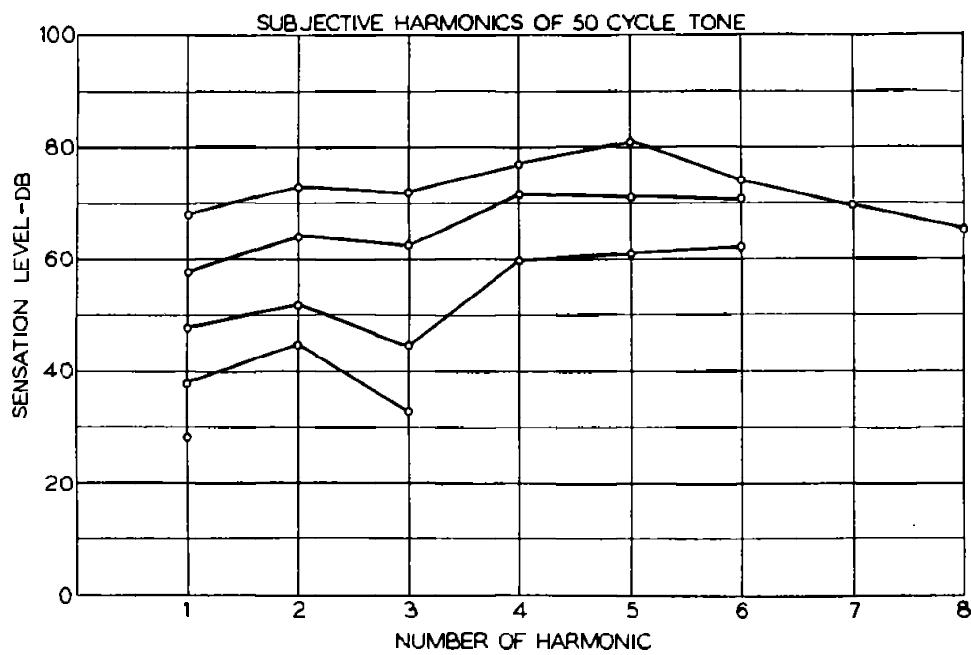


FIG. 3

from average curves. Due to this non-linearity of the middle ear transmitting mechanism the ratio of the amplitudes in the time-pressure pattern in the cochlea to the corresponding ones in the air in front of the ear drum will be much less for the high than for the low intensities. As the intensities increase the transmitting system becomes more and more overloaded and consequently the output diminishes comparatively.

For intensity levels below  $-6$  bels, the output is proportional to the input. For higher levels the output drops below the one to one ratio. If during the vibration, the displacement at the oval window is the same on either side of the neutral position, then only the odd harmonics

would be present. Since both even and odd harmonics are found the vibration must be unsymmetrical and a rectifying action must take place. Thus we have both dissymmetry and non-linearity in the process of transmission.

### MECHANICS OF THE MIDDLE EAR

To simplify the kind of motion taking place in the cochlea, we shall first consider the case when no subjective harmonics are present. Let a pure tone be transmitted through the middle ear to the fluid in front of the oval window in the cochlea. The question then is what will be the motion of the fluid in the inner ear and in particular what will be the motion of the basilar membrane where the nerve endings are located.

At the outset let me emphasize that no satisfactory solution of this problem has been made nor is a solution offered here. Wegel and Lane<sup>6</sup> in their paper on "The Auditory Masking of One Pure Tone by Another and Its Probable Relation to the Dynamics of the Inner Ear", made a good approach to the problem when they suggested an electrical analogue of the cochlea. However, they made no numerical calculations due to the uncertainty of the constants involved. In spite of the many objections which may be raised to the simple treatment given below still it probably helps one to better understand the action of the cochlea.

According to the Helmholtz resonance theory, the vibration communicated to the liquid from the stapes is transmitted a certain distance through the scala vestibuli depending upon the pitch of the tone, and then through the basilar membrane, back through the scala tympani to the round window. The nerve endings at the position where the vibration is transmitted through the basilar membrane are stimulated and send impulses to the auditory nerve which carries them to the brain.

Consider a small element of the basilar membrane of length  $\delta x$  and width  $w$  and located at a distance  $x$  from the oval window. The sinusoidal pressure applied at the oval window is transmitted through the liquid and causes this element to vibrate. The problem then is to determine the magnitude of its vibration.

If the variation in pressure is slow all surfaces of equal area in the cochlea will experience the same force at practically the same time. Consequently, the forces on either side of the element under consideration are equal and opposite and would therefore keep the element

stationary. As the pressure variation becomes more rapid, there will develop an excess pressure on the upper side which will cause the element to move. The force  $F$  due to this excess pressure will be

$$F = k_F w \cdot \delta x P \cos \omega t \quad (1)$$

where  $P$  is the pressure amplitude at the oval window and  $\omega/2\pi$  is the frequency of variation.

For the reasons given above we have introduced the parameter  $k_F$  which varies from unity at high frequencies to zero at very low ones.

The reactive forces opposed to the impressed force mentioned above are due mainly to the inertia and frictional resistance of the liquid moved and the stiffness of the basilar membrane at the point considered.

The effective mass is due almost entirely to the liquid on either side of the element. It seems reasonable to suppose that this effective mass  $m$  is given by

$$m = k_m w \cdot \delta x \cdot x. \quad (2)$$

Stated in words this means that the effective moving mass of liquid is proportional to the distance the element is away from the oval window and proportional to the amount of liquid in contact with the element. It is evident that this would be true for the simple type of motion depicted in Fig. 4, where a column of liquid reading from the oval window to the position  $x$  and back on the other side of the

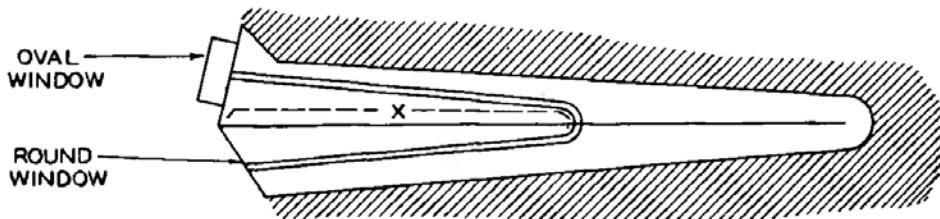


FIG. 4

membrane to the round window, and having a cross section  $w \cdot \delta x$ , moves to and fro as a unit. In this case  $k_m = 2$  since the density of the liquid is approximately unity. Although the motion cannot be as simple as that described yet the general patterns of the motion due to tones of different frequencies are probably somewhat similar and differ only in size. For these reasons one might expect Equation (2) to be approximately correct.

The stiffness  $S$  is principally due to the element itself, although of course there is some stiffness offered to the movement of the liquid

by the round window and the oval window and the bones attached to it. The stiffness factor is proportional to the number of cross fibres in the element  $\delta x$ , to their average tension  $T$ , and inversely proportional to their length  $w$ , that is the width of the membrane at the position  $x$ . Therefore

$$S = k_s \frac{\delta x T}{w}. \quad (3)$$

It is well known that such a system has the natural frequency  $f_0$  given by

$$\omega^2 = (2\pi f_0)^2 = \frac{S}{m} = \frac{k_s}{k_m} \frac{T}{w^2 x}. \quad (4)$$

This then is an equation which gives a relation between the position  $x$  on the basilar membrane for maximum stimulation and the frequency  $f_0$  of the tone impressed upon the ear.

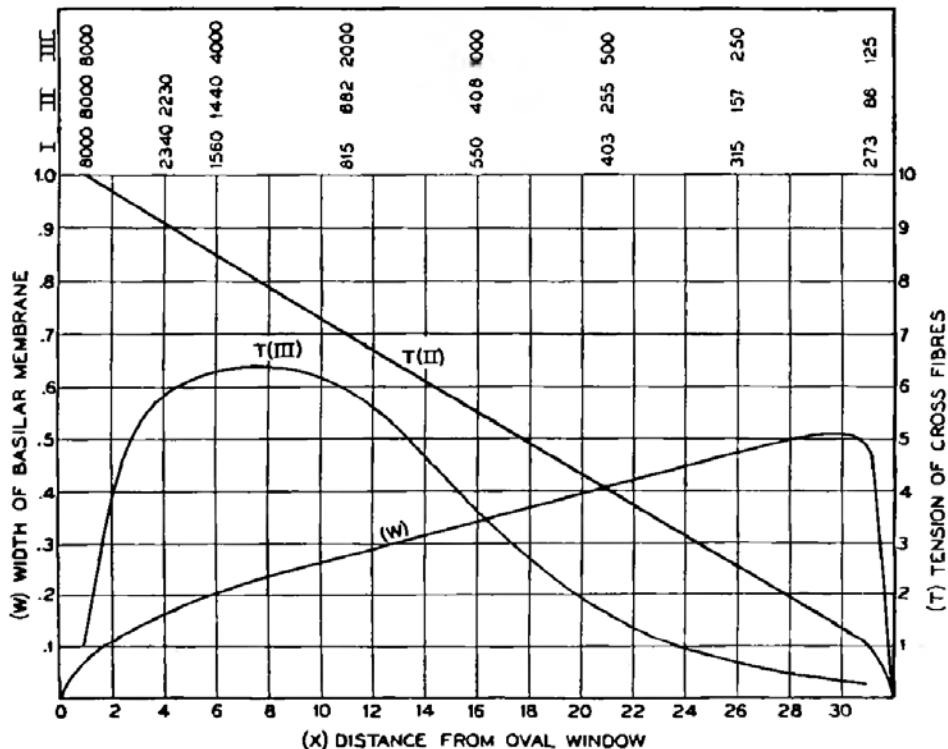


FIG. 5

Let us first consider the anatomical data in connection with equation (4) to see what they would predict concerning the positions of maximum stimulation for tones of different pitch.

The width  $w$  can be obtained fairly accurately from microscopic

cross-sections. Although there is considerable variation from sample to sample, the curve marked " $w$ " in Fig. 5, may be taken as representing an average result. It is seen that  $w$  varies from .08 to .5 as  $x$  varies from 1 to 31 mm.

The anatomical data indicate that the tension  $T$  of the cross fibres decreases as  $x$  increases but no quantitative data are available. Three cases are postulated and the calculated results are shown in Fig. 5. In Case I, it is assumed that the tension of the cross fibre is constant throughout the length of the membrane. In Case II it is assumed that the tension decreases uniformly with the ratio of 10 to 1 in going from the oval window to the helicotrema. In Case III the vibration of  $T$  is chosen so as to give a uniform distribution along the basilar membrane of the pitch covering six octaves. In all cases  $k_s/k_m$  was determined by assuming that for the resonator at  $x=1$  the resonant frequency was 8000 cycles per second.

It is seen from the anatomy of the cochlea that we would expect the position of the maximum stimulation for the high frequency to be near the oval window and move toward the helicotrema as the frequency is lowered. It is seen that the necessary range of resonance may be covered without making any radical assumptions regarding the variation of the tension of the cross fibres. Even though no variation in the tension occurs still the range of resonance is from 273 to 8000 cycles per second.

Most authors who have tried to make a simple analysis of this problem have assumed that the frequency of resonance is inversely proportional to the length  $w$  of the cross fibres, but this analysis makes it inversely proportional to  $w^2$ . This is because it was assumed that the effective mass due to the vibrating liquid which is associated with a small portion of the basilar membrane, decreases as  $w$  decreases. In other words, the effective vibrating column is just sufficiently wide to cover the basilar membrane. A ten-fold variation in  $T$  accounts for a variation in  $f_0$  from 86 to 8000 cycles per second. If we assume that the rods of Corti extend to values of  $x=.05$  cm. and that  $w$  varies as indicated, then this range will be extended to 16000 cycles per second. This is only true provided that the equation can be applied to this end condition, an assumption which is probably not true.

The experimental data on minimum perceptible differences in pitch<sup>6</sup> indicate that the distribution frequency along the basilar membrane

<sup>6</sup> This was pointed out by Wegel & Lane in the paper already cited.

corresponds more nearly to that of Case III. For this reason this simple relation between  $f_0$  and  $x$  will be used in the following discussions. Mathematically stated then, the resonant pitch

$$P_0 = \frac{14}{5} - \frac{x}{5} \quad (5)$$

or

$$\log_{10} f_0 = \frac{16 - x}{16.6} + 3 \quad (6)$$

which holds for values of  $x$  from 1 to 31. According to this view then, the positions of maximum response are distributed uniformly along the basilar membrane according to the pitch of the exciting tone and cover a range of six octaves from -3 to +3. Impressed tones having pitches above or below this range do not create a condition of resonance in the cochlea. This may be one of the reasons why the sensitivity drops off for tones having pitches beyond these limits.

Having established the relation between  $f_0$  and  $x$ , let us now consider the form of vibration of the basilar membrane. If we were dealing with separate stable resonators instead of mutually connected unstable ones, the problem would be simple. But as we have seen this is not the case. The stiffness of a given portion of the membrane remains essentially constant for impressed tones of different frequency but the effective mass vibrating with it and also the resistance  $r$  changes.

The subsidence of the vibration after the force stops is dependent upon the ratio of the resistance- $r$  to the effective mass  $m$ , that is, upon the damping factor which is given in bels per second by

$$\Delta = .434 \frac{r}{m} \quad (7)$$

It is difficult to make any accurate estimate of  $\Delta$  from the anatomy of the cochlea but the following consideration may help to understand the factors involved.

Consider a tube of liquid of length  $2x$  and having a cross  $w \cdot \delta x$  moving slowly to and fro in a larger tube filled with the same liquid. This is illustrated in cross section by Fig. 6. When the cross sections of such tubes are small the frictional force is proportional to the interacting surface namely  $2x \cdot 2(w + \delta x)$ , multiplied by the velocity, the factor of proportionality being  $\eta/d$  where  $\eta$  is the coefficient of viscosity, and  $d$  the effective distance of the surface from the containing walls.

The value of  $d$  is approximately that shown in Fig. 6. The damping factor for such a vibrating column when  $\delta x$  is small compared to  $w$ , will then be approximately

$$\Delta = .868 \frac{\eta}{\delta x \cdot d} . \quad (8)$$

It is seen that  $\Delta$  is independent of the length of the column but depends upon its thickness  $\delta x$ . If the thickness is just sufficient to cover a rod of Corti then  $\Delta = 30$  and if it is equal to  $w$  then  $\Delta = 570$  bels per second.

It would be expected that the value of  $\Delta$  for the case of the element of the basilar membrane considered above would be between these limits. Also this type of consideration leads one to suppose that  $\Delta$  is

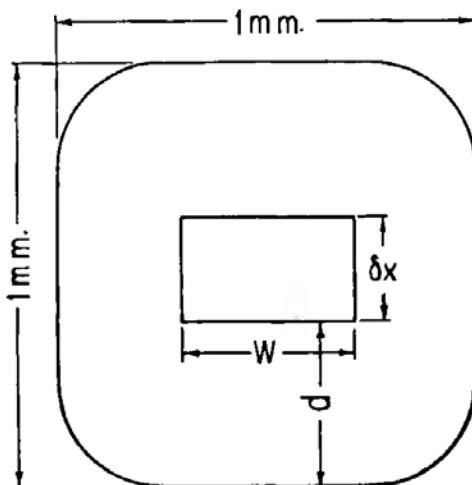


FIG. 6

independent of  $x$ , that is, it is the same for different positions along the basilar membrane.

Using the values given for  $F$ ,  $m$  and  $s$  we can now write down the expression for the displacement  $y$  of the element  $\delta x$  from its position of equilibrium

$$y = A \cos (\omega t - \Theta) \quad (9)$$

where the amplitude  $A$  is given by

$$A = \frac{1}{4\pi^2} \frac{kf}{k_m} \frac{P}{x[(f_0^2 - f^2) + (.37f\Delta)^2]^{1/2}} \quad (10)$$

and the phase  $\Theta$  by

$$\tan \Theta = .37\Delta \frac{f}{f_0^2 - f^2} . \quad (11)$$

It was seen that factor  $k_f$  varied from zero for very low frequencies to unity for high frequencies and that  $k_m$  had a value of approximately 2 at least if the motion were similar to that indicated in Fig. 6. Taking the value of  $k_f$  equal to unity,  $k_m$  equal to 2, and  $\Delta = 70$  and  $P = 1$  bar, calculations of  $A$  and  $\Theta$  were made. They are shown in Table I. The relation between  $f_e$  and  $x$  used was that given in equation (6).

TABLE I

$f=75$			$f=125$			$f=250$			$f=500$		
$x$	$A \times 10^8$	$\Theta$	$x$	$A \times 10^8$	$\Theta$	$x$	$A \times 10^8$	$\Theta$	$x$	$A \times 10^8$	$\Theta$
10	.2	0	10	.2	0°	10	.2	0	10	.2	0
20	2	0	20	2	1°	20	3	2°	15	.8	1°
25	7	2°	25	8	8°	25	30	27°	20	8	9
30	28	7°	30	71	33°	25.8	67	62°	20.5	16	19°
31	40	11°	31	126	90°	26	75	90°	21	46	90°
						27	28	157°	21.5	17	158°
						31	8	172°	22	9	168°
									25	3	176°
									31		

$f=1000$			$f=2000$			$f=4000$		
$x$	$A \times 10^8$	$\Theta$	$x$	$A \times 10^8$	$\Theta$	$x$	$A \times 10^8$	$\Theta$
10	.3	0°	5	.2	0°	5	.5	1°
15	2.6	5°	10	1.0	2°	5.5	1.0	2°
15.5	5.4	10°	10.5	2.0	5°	5.8	2.3	6°
16.0	30.4	90°	11.0	22	90°	6.0	20	90°
16.5	5.8	169°	11.5	2.3	174°	6.2	2.4	173°
17.0	3.1	174°	12.0	1.1	177	6.5	1.0	177°
20.0	.9	178°	15.0	.3	179°	7.0	.7	178°
31	.4	178°	31	.1	179°	20	.04	180°

The values of  $A$  are given in  $10^{-8}$  cm which is the order of magnitude for the diameter of the molecules. If values of velocity or acceleration are calculated for any one frequency the relative values will be the same as those shown in the table. Even the bending of the membrane will vary sinusoidally with amplitudes which are not greatly different from those shown. This suggests that the stimulation pattern for a tone without harmonics (either objective or subjective) will be the same as the amplitude pattern.

It will be seen that the resonant peaks are very much sharper for the high frequencies than for the low ones, a fact which is in agreement with the data on masking as will be seen.

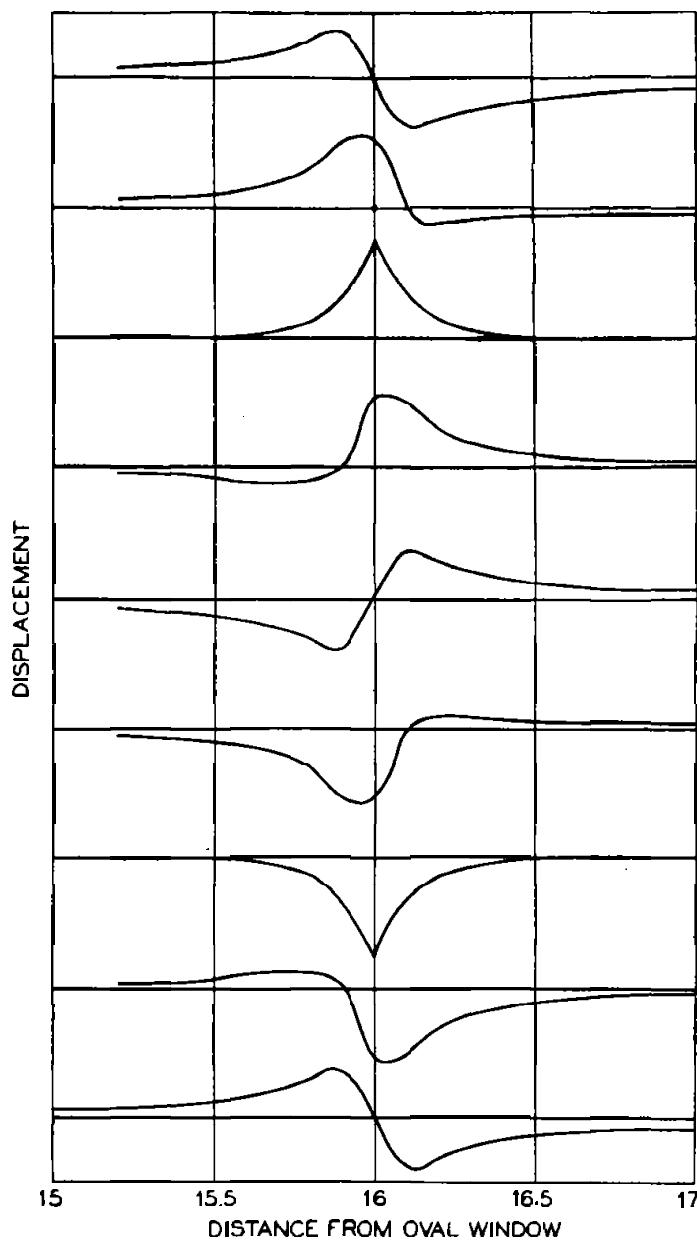


FIG. 7

It is important to notice that, according to these assumptions and most any other reasonable assumption one can make, the part of the membrane next to the oval window leads in the phase of vibration, so that if motion pictures were taken with exposures  $T/8$  seconds apart where  $T$  is the period for one complete cycle, we would get a series of

pictures like those shown in Fig. 7 which are plotted from the results in Table I for the case of 1000 cycles.

Bekesy<sup>7</sup> has pointed out that such a motion of the membrane would tend to produce whirls in the liquid above the position of maximum stimulation. When the exciting tones were loud, he observed such whirls both in his models of the ear and in dissected human ears. He considers that it is the steady pressure against the membrane produced by this whirl that causes the nerve stimulation. He points out that such a whirl would cause a movement of the liquid in the semi-circular canals so as to effect the sense of balance and cause an observer exposed to such a loud tone of high pitch to tilt his head toward the ear receiving the sound. He claims that such an effect has been noted a sufficient number of times to substantiate this point of view. It seems doubtful, however, that such a whirl would be set up except for very loud tones. It also seems difficult to explain masking data on any assumption other than that the stimulation of the nerve endings is due to the vibration of the basilar membrane which moves in some way similar to that described. The mechanism of this stimulation will be dealt with later.

TABLE II

<i>f</i>	Relative Reciprocal Threshold Pressures	Relative Displacements	Relative Velocities	Relative Accelerations
75	.004	1.22	.09	.007
125	.024	4.14	.52	.065
250	.125	2.47	.62	.17
500	.500	1.52	.76	.38
1000	1.000	1.00	1.00	1.00
2000	1.250	.73	1.47	2.84
4000	1.250	.62	2.48	9.92
8000	.180	.64 } .027 }	5.12 } .22 }	40.96 } 1.72 }
12000	.050	.0036	.043	.52

In Table II the relative values of the maximum displacement, velocity, and acceleration of impressed tones of various frequencies *f*, as calculated above, and also the reciprocal values of the pressure change in the air necessary to make the tone just audible are given. The second value at 8000 was calculated on the assumption that the nerve

<sup>7</sup> Phys. Zeit. Vol. 29, pp. 793-810, Nov. 15, 1928.

endings stop at  $x = 1.5$  instead of at  $x = 1.0$  mm. It is seen that for frequencies above 1000 cycles the threshold values agree better with the maximum velocity values and for frequencies below 1000 cycles better with the maximum acceleration values. But these calculations were made on the assumption that  $P$ , the pressure at the oval window, was constant for tones of different frequencies. As was pointed out above, the results of impedance measurements<sup>8</sup> upon the ear indicate that for constant intensity tones in the air having frequencies below 1000 cycles, a constant amplitude rather than a constant velocity of the ear drum is produced. Consequently, the value of  $P$  will be proportional to the frequency for such tones in this frequency range, assuming, of course, that no other change takes place in the process of transmission through the middle ear. Introducing this new factor into the calculation, the velocity values in this frequency range will be the same as those given under the heading "Relative Accelerations" in Table II. With this correction then, the velocity values agree with as good an accuracy as could possibly be expected from such an analysis with the reciprocal threshold values throughout the entire range. The factor  $k_f$  becomes smaller for the low frequencies which likewise tends to make the velocities smaller, especially those of very low frequencies. Although this evidence is far from conclusive, it is evidence for supposing that equal stimulations will be produced at different parts of the basilar membrane which have equal velocities even though the frequencies are different. For obvious reasons it would not be expected that Equations (10) and (11) would give values which would agree with the stimulation space pattern except for small distances on either side of the maximum stimulation position.

#### STIMULATION PATTERNS DETERMINED FROM MASKING DATA

These space patterns for the entire length of the membrane, however, can be obtained from the masking data. Such data taken on my left ear by the method described by Wegel and Lane<sup>9</sup> are given in Figs. 8 and 9 for stimulating tones having frequencies 75, 125, 250, 500 and 1000. The intensity level of the tones in the ear were adjusted to be equal to -20 db. To do this the sensation levels were made equal to 27, 40, 54, 65 and 72 db as indicated in the figures. The vertical lines represent the sensation levels of the subjective harmonics as deter-

<sup>8</sup> See Tröger's results page 34 of paper already cited.

<sup>9</sup> Phys. Rev. 19, 492, May, 1922.

mined by the "best beat" method. The ordinates of the curves give the masking, that is, the number of decibels that the threshold of audibility is shifted due to the presence of the stimulating tone. It is interesting to note that under these conditions the same number of

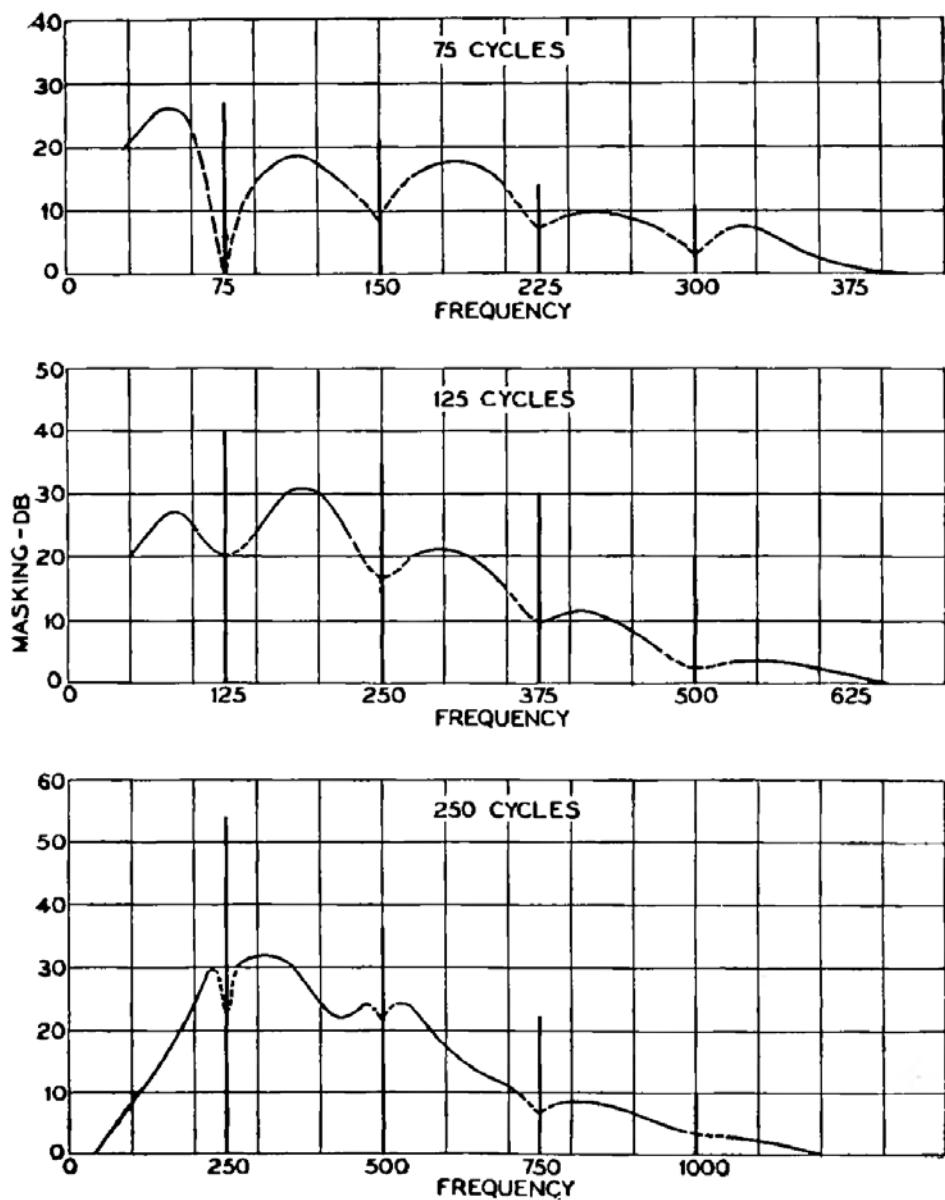


FIG. 8

audible subjective harmonics are created by each of the five stimulating tones and with approximately the same intensity levels. For example, the intensity levels of the second harmonic in each case are 42, 41, 49, 34 and 42 decibels below zero level. Similarly, the third harmonics are

56, 52, 67, 60, 53 and the fourth harmonics are 67, 75, 82, 85 and 86 decibels below zero level.

For the 75 cycle tone and for tones of lower pitch subjective harmonics could be detected as long as the tone could be heard. For

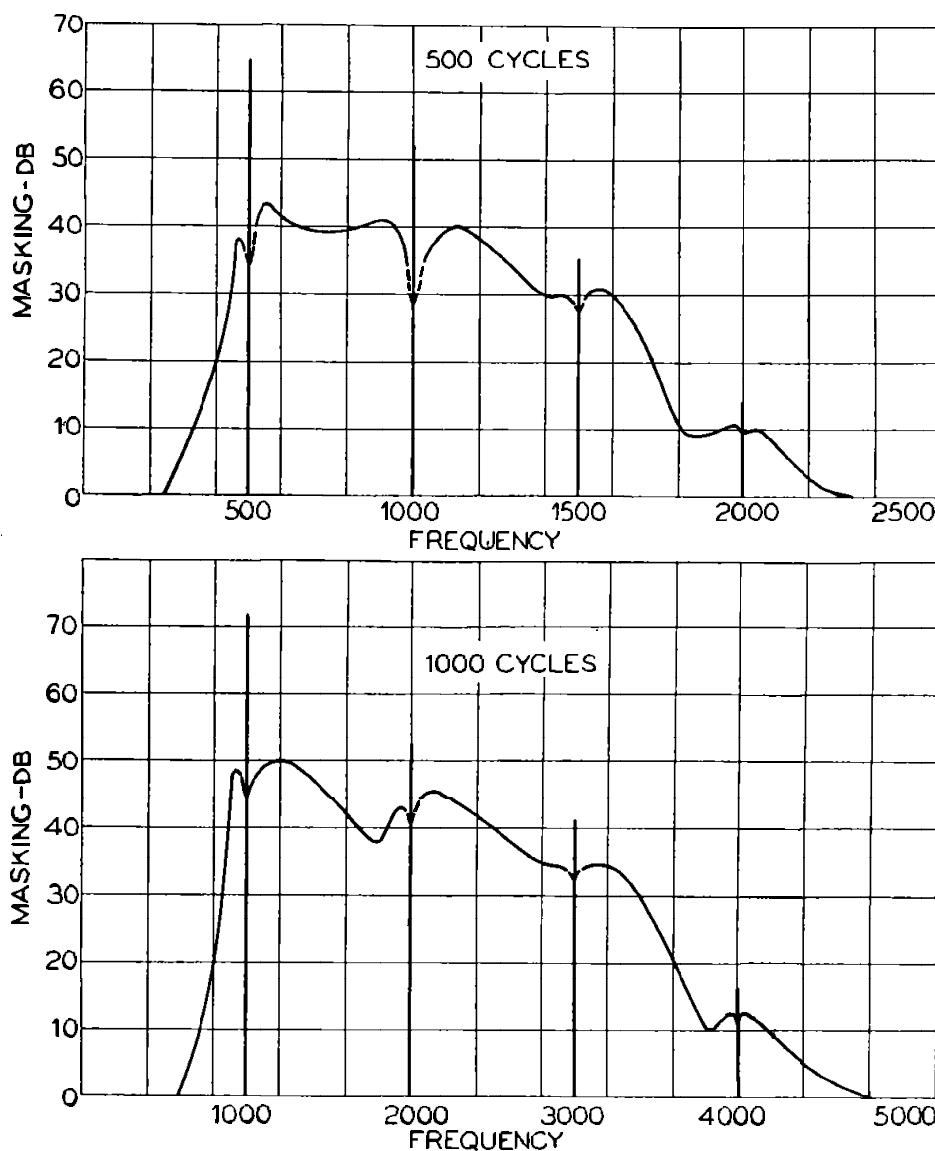


FIG. 9

example, if tones having frequencies of 50 and 103 are applied to the ear simultaneously, then for any intensity of the 50 cycle tone that can be heard, an intensity of the 103 cycle tone can be found so that the sensation of 3 beats per second will be produced. This fact makes it difficult to use masking data for determining the form of vibration of

the basilar membrane for a pure low pitched tone being impressed at the oval window of the cochlea.

If now we use the relation between  $f$  and  $x$  given by Equation (10), we can estimate stimulation level curves from these masking curves. The level for the position of the harmonics will be directly that given in Figs. 8 and 9. For intermediate positions where only masking takes place, the procedure is more uncertain. If we assume that the masked tone is just perceptible when it is at a level corresponding to that necessary when the masking and masked tone have the same frequency

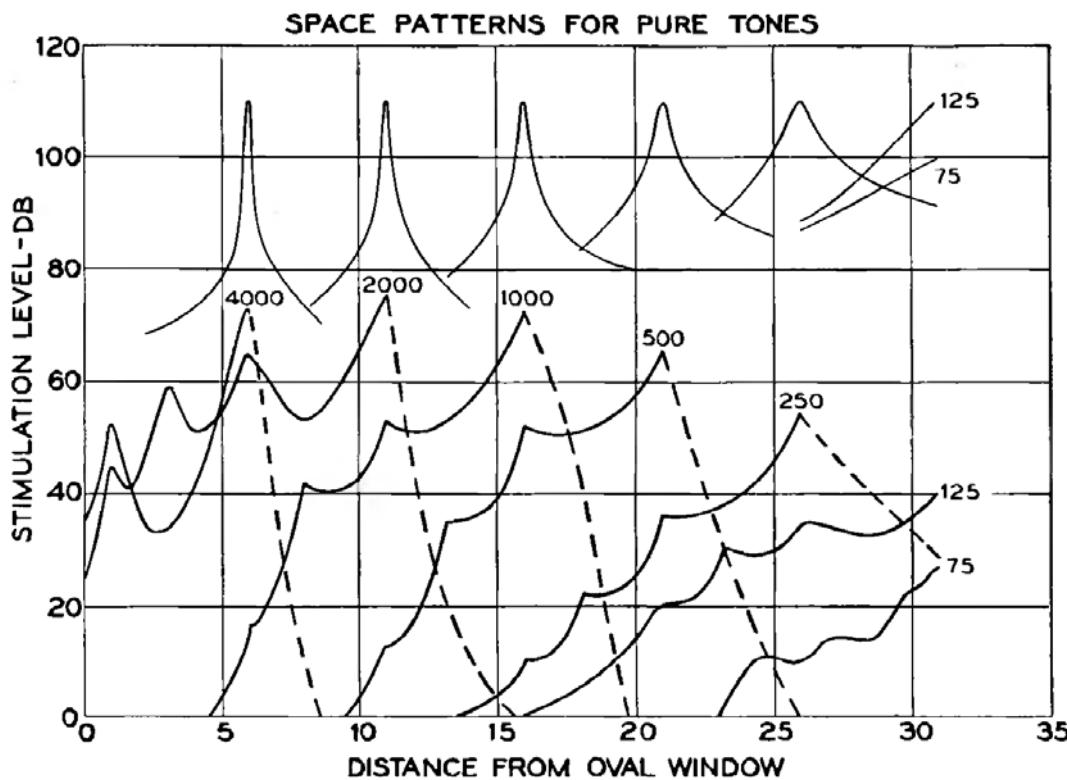


FIG. 10

then the data on minimum perceptible differences in intensity can be applied. These data indicate that when the masking tone has a greater sensation level than 40 db then the masked tone is always approximately at a level which is 20 db lower than the level of the masking tone. As the sensation level of the masking tone decreases below 40 db the difference between the levels of the masking tone and the masked tone decreases from 20 db to zero. In this way the curves given in Fig. 10 were obtained. The right side of each curve is shown dotted so that the curve might be followed more easily. The curves indicated by the light lines at the top of the figure are the calculated ones taken from Table I,

the amplitude values being transformed to decibels and the peak values adjusted to be at the same level. It is seen that for a range of 3 or 4 mm. along the basilar membrane, these curves have the same general shape as the masking data curves. For greater range the latter curves fall off much more rapidly than the simple resonance curves. The curves for 2000 cycles and 4000 cycles were taken from data obtained by Mr. F. H. Graham.

It is seen then that when each of the five tones considered above is impressed upon the ear, space patterns similar to those shown in Fig. 10

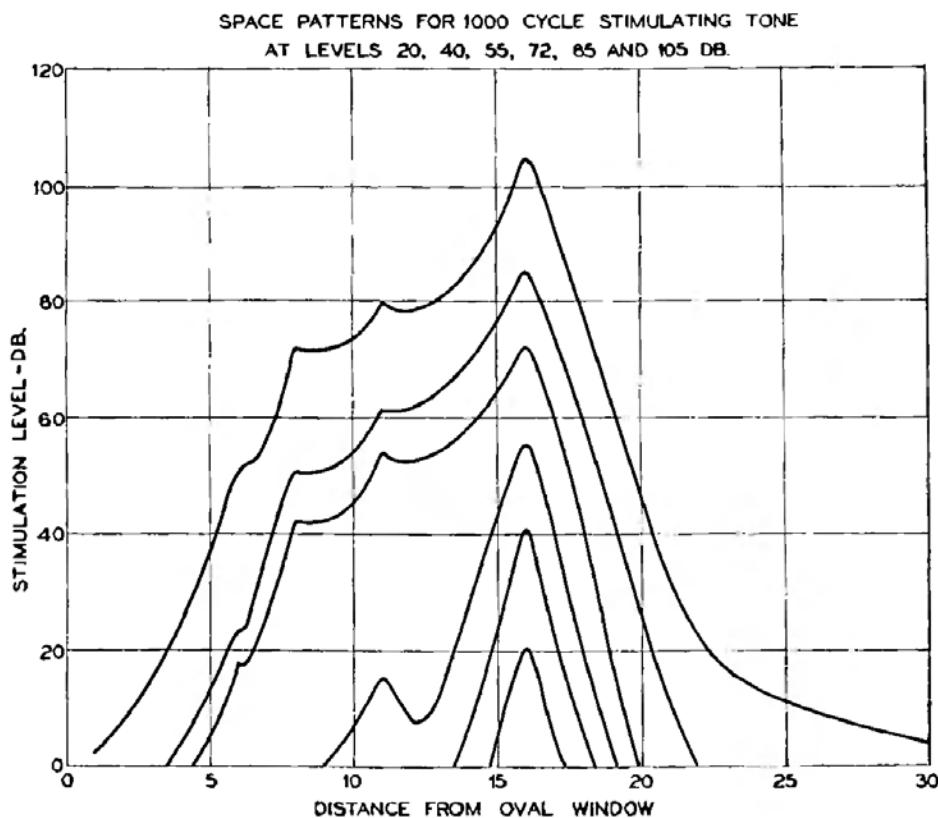


FIG. 11

are produced corresponding to the sinusoidal time pattern in the air.

Similarly, for every sound reaching the ear a space stimulation pattern corresponding to the time pattern of the sound in the air is produced on the basilar membrane. This space pattern is transferred to the brain by a nerve cable containing about 3000 individual nerve fibres. This space pattern is dependent both upon the kind of sound and also the intensity with which it is applied to the ear. This is illustrated by the stimulation patterns shown in Fig. 11 for a 1000 cycle tone impressed with the various intensities indicated by the peaks of the

curves. These curves were estimated from data on best beats and masking as described above.

#### MECHANISM OF NERVE STIMULATION

Before considering how these stimulation patterns are related to loudness, let us consider more closely the mechanism of stimulating the nerves. Some of the nerve endings will be very sensitive while others will be very insensitive; others will have degrees of sensitivity scattered between these extremes. Let us take the stimulation level at a point on the basilar membrane as the number of bels that the velocity at that point is above the agitation of the membrane which would just stimulate the most sensitive fibre in a group situated along one millimeter of length of basilar membrane. Then if the nerve endings in the ear have properties similar to those in the eye the distribution of threshold values will be as follows. Let  $z$  be the fraction of fibres which have threshold values below the stimulation level  $\beta$ , then

$$\beta - \beta_0 = \log \frac{z^2}{1-z}. \quad (12)$$

The constant  $\beta_0$  is the value of the right-hand member at the threshold. As shown by Hecht<sup>10</sup> this relation holds for the nerve endings in the eye. The constant  $\beta_0$  depends upon the value of  $z$  at the threshold.

As the stimulation level goes above the threshold of a nerve ending the firing of nerve impulses from that ending becomes more rapid. Let us consider this mechanism for a moment. We will assume that there are two opposing operations in the process of stimulating the nerves. The first, which may be called the active process, is dependent upon the magnitude of the motion and tends to create either a substance or energy which acts like a catalytic agent and produces a nerve discharge. The second, which may be called the reactive process, tends to destroy or annul the active process. For example, let us call the thing being created the discharger which may be the energy of agitation of the element of the nerve ending or a fluid produced by some chemical process or some kind of electrical accumulation. The active process will tend to create this discharger at a rate depending upon the stimulating force and the reactive process will destroy the discharger at a rate depending upon the amount present. If the discharger is the amount of

<sup>10</sup> Light Sensitivity of Animals. Proceedings of the Michelson Mtg. of the Optical Society of Amer., Nov., 1928.

energy of vibration, then the action of frictional forces furnishes the reactive process. Let us call the amount of discharger present  $q$ . As a first approximation it will be assumed that the rate at which  $q$  is created is proportional to the energy<sup>11</sup> of agitation and the rate at which it is destroyed is proportional to  $q$ . Formulating these assumptions into a differential equation

$$\frac{dq}{dt} = av^2 - bq \quad (13)$$

where  $a$  and  $b$  are quantities independent of  $x$  or  $t$  and  $v$  is the velocity of the membrane where the nerve ending is situated.

For a periodic force such as that produced by a pure tone

$$v = V \cos \omega t. \quad (14)$$

Substituting this in Equation (13) and solving under the conditions

$$q = 0 \text{ when } t = 0$$

$$q = \frac{aV^2}{2b} \left( 1 - e^{-bt} + \frac{1}{\sqrt{1 + \left(\frac{2\omega}{b}\right)^2}} [\cos(2\omega t - \Theta) - e^{-bt} \cos \Theta] \right) \quad (15)$$

where  $\tan \Theta = 2\omega/b$ . If  $b$  is small compared to  $\omega$  then

$$q = \frac{aV^2}{2b} \left( 1 - e^{-bt} - \frac{b}{2\omega} \sin 2\omega t \right). \quad (16)$$

The amount of discharger present is proportional to the energy of agitation, inversely proportional to  $b$ , and directly proportional to the time factor in the parentheses. This process continues until the threshold value of the nerve ending is reached when a new process starts; a nerve discharge then takes place. While this second process is going on the equilibrium is upset so that the creation of the discharger stops. This period of the stopping of the first process is the reaction time  $\tau$  of the nerve ending. After the reaction interval, the active process starts again. At the high intensities the number of discharges is governed mostly by the reaction time while at the low intensities mainly by the time for the discharger to accumulate.

<sup>11</sup> This assumption is arbitrary and results in predicting that the double frequency will occur in the brain time pattern. This is used only for illustrative purposes. Other assumptions can be made which will result in leaving the fundamental and its harmonics in the time pattern of the brain.

To illustrate the form of this equation, the graph of the time function in the brackets is shown in Fig. 12 when  $b = 125$  and the frequencies of 125 and 500 cycles per second. Any fibre may obtain the requisite amount of  $q$  for a discharge either by having a large velocity  $V$  and a small time  $t$  or a small velocity  $V$  and long time  $t$ . Consequently, for any given value of  $V$  the sensitive fibres will require only a short time to discharge and the insensitive ones a long time. It is seen that for the 500 cycle tone  $q$  increases in the interval .001 to .003 second about 10 fold while in the same length interval .003 to .005 it remains practically constant. It is evident then that the number of discharges coming from a stimulated patch of nerve fibres which take place in the first interval

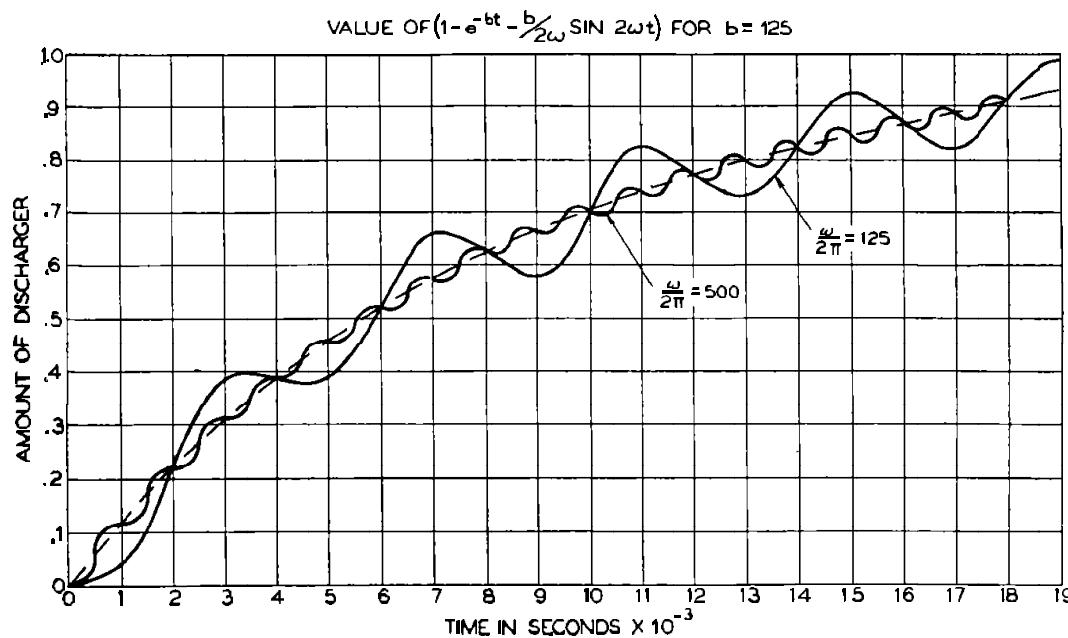


FIG. 12

will be very much greater than in the second interval. Consequently, the bombardment of nerve discharges reaching the brain will have times of maximum intensity which are separated by intervals equal to the half period of the oscillating body producing the stimulating sound. In other words, a time pattern of the changes taking place in the medium transmitting the sound wave is transmitted to the brain. In passing to the brain it follows the nerve tract shown in Fig. 13.

Although the nerve fibres are interrupted by synapse as indicated, where a restimulation is necessary, it seems reasonable to expect some sort of correlation between the rate that the nerve discharges are being received and the time pattern of the original wave. As the frequency

## AUDITORY TRACT

According to A. R. Rasmussen, Professor of Neurology at the University of Minnesota

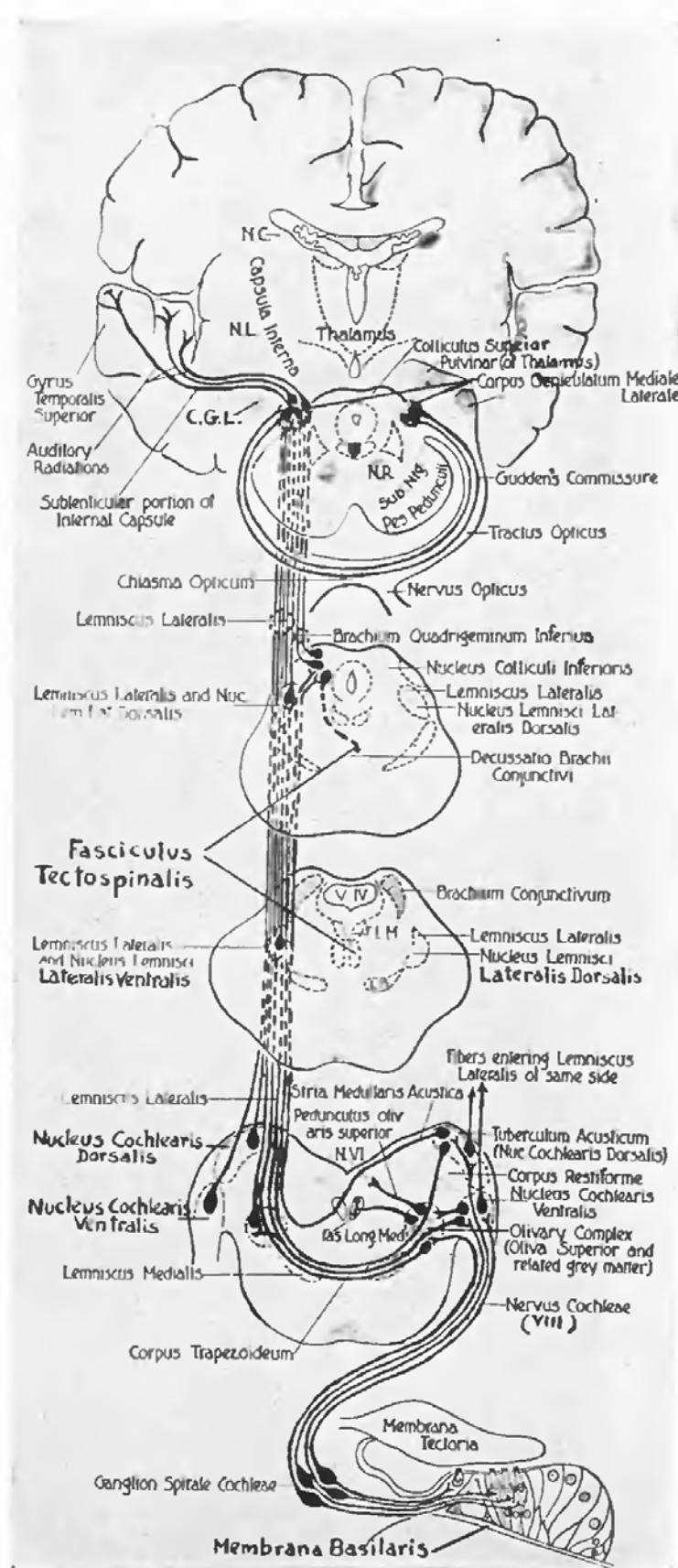


FIG. 13

*Note:* You are looking down on transverse sections through various levels of the brain except in case of the upper part of the upper level (see below). The dorsal side of the brain is up.

Only one lateral lemniscus is carried up from level to level. If the fibers from the cochlear nuclei of the right-hand side were carried up they would follow a course similar to that taken by the fibers that cross to the left-hand side. The brain is bilaterally symmetrical.

The lower level, which is through the extreme lower part of the pons, is slightly oblique, being a little lower dorsally than ventrally so as to intersect the dorsal cochlear nucleus.

The second level (from below, not counting the drawing of the organ of Corti) is through the middle of the pons. The third level is through the upper part of the pons ventrally and lower part of the mid-brain dorsally. The lower part of the upper level shown is a cross-section through the upper part of the mid-brain. In order to show the position of the auditory radiations, it was necessary to resort to a vertical section through the cerebral hemispheres. This vertical section is so placed as to intersect the cross-section level of the upper mid-brain in the region of the medial geniculate body.

of the stimulating tone becomes higher the time pattern becomes less distinct since the ratio  $b/2\omega$  becomes smaller. The case for 500 cycles is also shown in Fig. 12. The variation from the exponential lines for a 5000 cycle tone would be 1/10 that shown for the 500 cycle tone.

For a large value of  $V$  the nerve will discharge in a very short time and for small values it will take a longer time and may not discharge at all. Let  $V_1$  be a value corresponding to a stimulation level  $\beta$  such that  $q$  will be at the threshold for a certain fibre when acted upon for a time long enough that  $e^{-bt}$  is negligible compared to unity, that is, so that the time factor in Equation (16) has reached its maximum value of  $(1+b/2\omega)$ . Let  $V_2$  be any other velocity above  $V_1$  corresponding to a level  $\alpha$  which will give the same value of  $q$  in a time  $t$ . Then

$$V_1^2 \left( 1 + \frac{b}{2\omega} \right) = V_2^2 \left( 1 - e^{-bt} - \frac{b}{2\omega} \sin 2\omega t \right) \quad (17)$$

$$\alpha - \beta = \log \frac{1 + \frac{b}{2\omega}}{1 - e^{-bt} - \frac{b}{2\omega} \sin \omega t} \cdot \log \frac{b}{2\omega} \quad (18)$$

The value  $t$  in this expression is the time necessary for a nerve ending to be acted upon at a level which is  $\alpha - \beta$  bels above its threshold level before it starts a nerve discharge. As mentioned above, the reaction time  $\tau$  is the time necessary for the nerve to discharge and recuperate, thus being ready again for another stimulation. Hence the time  $t + \tau$  is time interval before the nerve is ready to start the process over again. At the end of this time interval the phase of vibration will not in general be the same as in the beginning. However, on the average, we can neglect the term  $b/2\omega \sin 2\omega t$  when determining at each level the average rate of firing since it simply has the effect of making the time interval oscillate on either side of the value obtained when this term is omitted. The rate so determined can then be applied to all frequencies and it depends only upon the stimulation level. With this understanding the average rate of firing from a nerve which is stimulated at a level  $(\alpha - \beta)$  bels above its threshold is given by

$$\tau = \frac{1}{t + \tau}. \quad (19)$$

The fraction of threshold values lying between the levels  $\beta$  and  $\beta + d\beta$  is  $(dz/d\beta)d\beta$ . Consequently, if  $n$  is number of fibres per mm. the number of nerve discharges per second coming from a small length  $dx$  of basilar membrane will be given by the expression

$$ndx \int_0^\alpha r \cdot \frac{dz}{d\beta} d\beta. \quad (20)$$

For convenience let

$$S(\alpha) = \int_0^\alpha r \cdot \frac{dz}{d\beta} d\beta. \quad (21)$$

The value of  $dz/d\beta$  can be obtained from Equation (12) and the value of  $r$  from (18) and (19). The total rate of discharge  $R$  going to the brain is then

$$R = \int_0^{\alpha} nS(\alpha)dx. \quad (22)$$

Although the values of these integrals cannot be obtained analytically, they can be readily evaluated by graphical methods. To find the value of  $S(\alpha)$  change the variable from  $\beta$  to  $z$  then

$$S(\alpha) = \int_{z_0}^z \cdot r \cdot dz \quad (23)$$

where  $z_0$  corresponds to the threshold value, that is, the fraction of fibres that must be excited to produce a sensation. Taking  $\tau = .002$  second and  $b = 125$ , and  $z_0 = 10^{-4}$  values of  $S(\alpha)$  were calculated as shown in Fig. 14. The ordinates give values of  $S(\alpha)$  which is a number that when multiplied by the number of nerve endings at the stimulation level  $\alpha$  will give the total discharges per second from such nerve endings. There are about 1000 rods of Corti per millimeter along the basilar membrane. Then, according to this calculation, there would be 10 discharges per second from a millimeter at the threshold or zero stimulation level and 500,000 discharges per second when stimulated for its maximum discharge rate.

#### COMPARISON OF THE SPACE-TIME PATTERN THEORY WITH THE EXPERIMENTAL FACTS OF AUDITION

It was seen that variation in ear sensitivity with frequency was satisfactorily explained by the variation of the impedance of the ear,

and variation of velocity of basilar membrane due to the mechanical action of the cochlea.

There is no doubt that loudness is closely associated with the total discharges reaching the brain. It is seen by Fig. 14 that the possible range of loudness from a single mm. of membrane vibrating at different amplitudes is 11 or 12 bels. If we add to the one millimeter the possible 30 millimeters, the total impulses will be multiplied by 30 which is equivalent to increasing the level 2 or 3 bels in the high intensity regions. It is thus seen that this mechanism provides for ample range of the

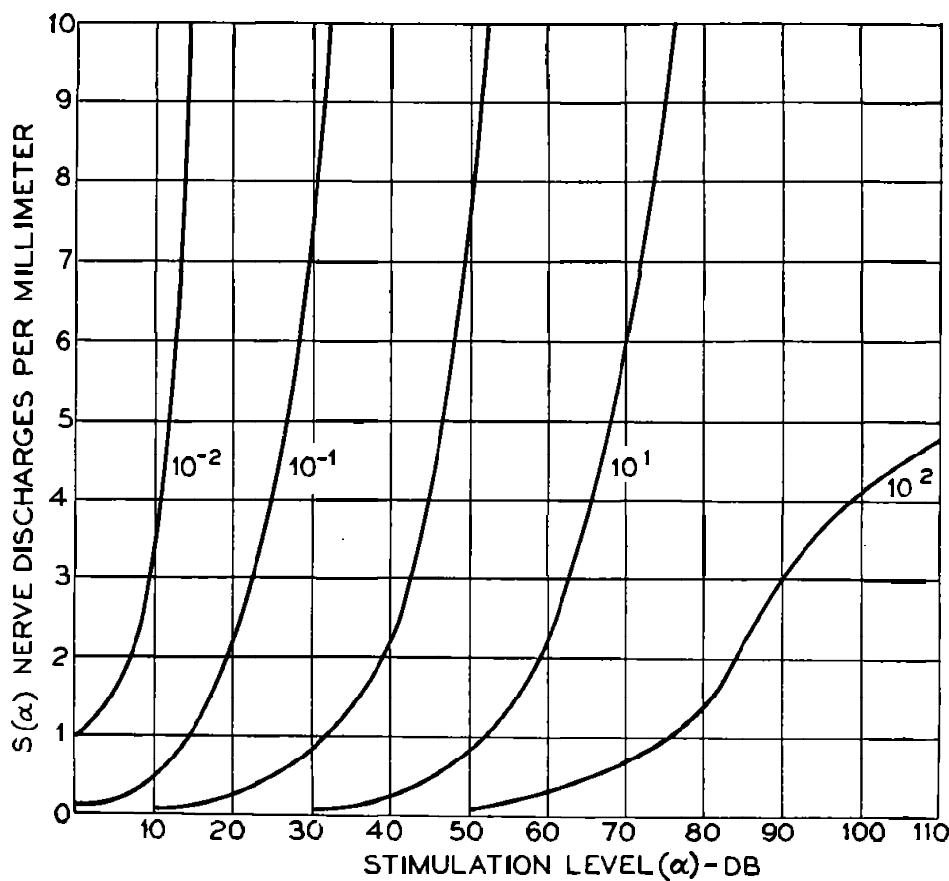


FIG. 14

loudness magnitudes used in audition. When a tone is gradually increased in intensity from the threshold, it first stimulates a small patch; this grows in length and then other patches corresponding to the subjective harmonics are started until finally some nerve fibres along the entire length are stimulated. If the membrane vibrated the same for all tones, similar for example to the diaphragm of the Wente condenser transmitter, then the loudness for all tones should be the same

for the same sensation level which is contrary to the observed facts. It is the difference in form of vibration of the basilar membrane that accounts for the difference in loudness of tones having the same sensation level.

The functions calculated above have been based on the assumption that as the intensity of stimulation increases the rate of discharging the nerve also increases until the limit of 500 per second is reached. Experiments on nerves have shown that as the nerve is continually stimulated it is fatigued, after which the maximum rate is very much reduced. Consequently, if a tone of given frequency stimulates the ear for some time then immediately after such stimulation the loudness of tones in the same frequency region will be reduced. This effect has been recently measured by Bekesy.<sup>12</sup> His results indicate that changes in loudness as much as 30 db occur immediately after strong stimulations.

Similarly, the threshold of audibility should be shifted if taken immediately after such strong stimulations. Measurements in our laboratory, which will soon be published, indicate that especially for low pitched tones such shifts amount to as much as 25 db and some shifts take place as long as 2 minutes after the stimulating tone is stopped. So the relations developed above can be applied only when the nerves are not fatigued. Both of these effects show that the changes are greatest for the tones near the pitch of the stimulating tone and decrease to zero for tones far removed from the pitch of the stimulating tone. This then is additional evidence that certain regions only along the basilar membrane are stimulated by a simple tone.

TABLE III

Frequency	Sensation Level	<i>R</i>	Calculated Loudness	Observed Loudness			
				<i>F</i>	<i>Ror</i>	<i>Re</i>	<i>G</i>
75	27 db	1,000	35	36	50	40	33
125	40	10,000	50	50	57		
250	54	40,000	61	60	62		
500	65	108,000	69	72	66		
2000	75	250,000	77	79	76		
4000	73	140,000	71	70			

It has been seen that due to the introduction of more subjective harmonics and also due to the form of vibration having less sharp resonant

<sup>12</sup> Phys. Zeit. Vol. 30, No. 21, Nov., 1929.

peaks, the region of maximum stimulation is much broader for tones of low frequency than for those of high frequency when both have the same sensation level. For this reason, these low tones will send more nerve impulses to the brain and consequently will be louder.

The values of  $R$  of Equation (22) were calculated for the seven tones shown in Fig. 10, taking  $n$  equal to 1000 per millimeter. The results are shown in the third column of Table III. Similarly, the values of  $R$  for a 1000 cycle tone at various levels were calculated from the curves in Fig. 11. The results are given by the curves in Fig. 15.

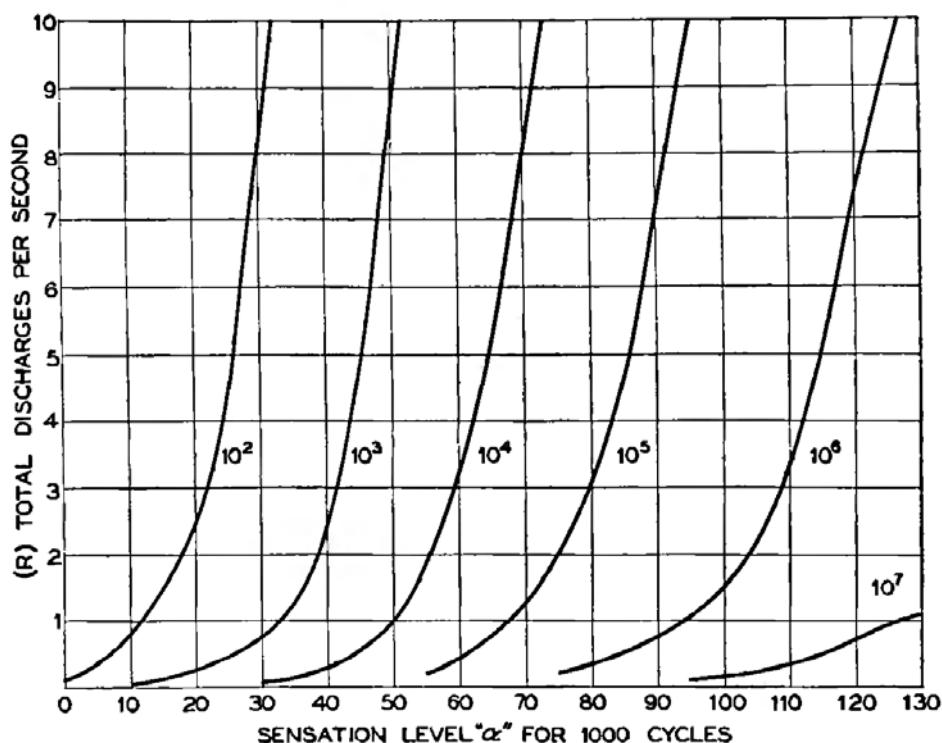


FIG. 15

The loudness of a sound is defined<sup>13</sup> as the sensation level of a 1000 cycle tone which is adjusted to sound equally loud. Consequently, the loudness numbers of each of the tones whose frequencies are shown in column (1) of Table III can be obtained by comparing the values of  $R$  with those in Fig. 15 and noting the corresponding sensation level. The values in the fourth column were obtained in this way. Inasmuch as all the data for constructing these curves were based upon measurements on my left ear, some loudness measurements were also made with this

<sup>13</sup> See Book "Speech & Hearing" page 226.

ear. The level of a 1000 cycle tone was adjusted so that it sounded equally loud to the six tones all introduced at the common intensity level of  $-20\text{ db}$ . The observer listened to the two tones alternately so that the fatigue effect should be eliminated. The observed results are shown in the fifth column. Inasmuch as these values were somewhat different than those reported by Kingsbury, I thought my judgment may have been influenced by knowing what to expect, so I asked another observer not familiar with the subject to make similar balances. His results are shown in the sixth column. Inasmuch as the difference was rather large for the 75 cycle tone, two other observers were asked to make balances for this tone. Their results are shown in the seventh and eighth columns. The agreement between the observed and calculated values seems to agree within the limits of observational error although the observed values tend to be higher for the two tones of lowest pitch. It is possible that the phenomenon of "volume" influences the observer to judge the low tones louder than otherwise would be the case. This may be one of the elements that enters into a judgment of equal loudness.

Let us now see how the differential sensitivity of the ear for pitch and intensity fits into the picture. These two effects are closely related for in each case it is a small change of the vibration form of the basilar membrane that can be detected, at least for the higher frequencies. First consider small changes in pitch.

Since the peaks are sharper for the high pitched tones, it would be expected that smaller changes in pitch could be detected for these tones which is in accordance with the facts. It would be expected also that smaller differences could be detected at the high intensities for under these conditions the several peaks due to the subjective harmonics will aid in making such detections. This also agrees with the observed facts. The smallest pitch change that can be detected is about .25 centi-octaves which corresponds to a shift in the stimulation peak of about 1/100 millimeter covering about 10 nerve endings. So this effect is not inconsistent with the anatomy of the ear.

The smallest intensity level change that can be detected is about .25  $\text{db}$ . This corresponds to about a 5% change in  $R$ . There seems to be no reason why this should or should not change with intensity level or pitch unless it is a change in the space pattern that is detected rather than a change in  $R$ . When two tones of the same frequency are compared it is probably a change in the space pattern that enables one to

detect the difference in intensity. If the change in form of the space pattern is the dominating influence for such small changes in intensity, then perceptible changes in intensity should be influenced by the same factors as those mentioned for perceptible pitch changes. This latter view seems to agree with the experimental facts concerning differential sensitivity for intensity.

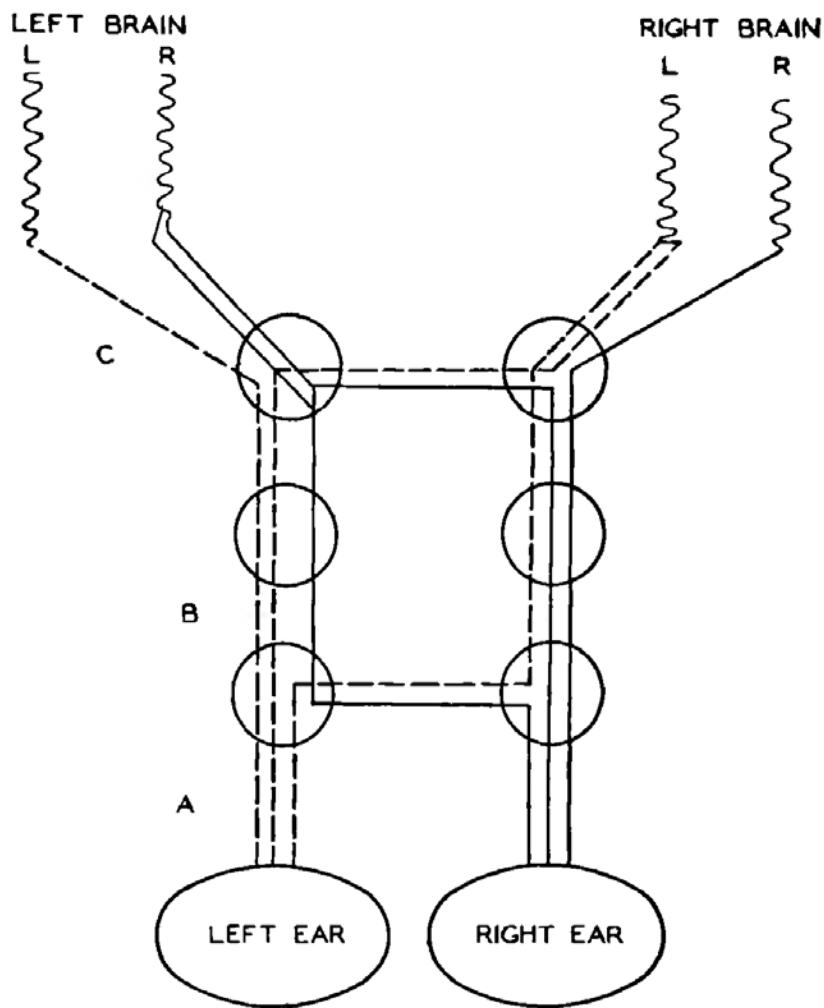


FIG. 16

It also agrees with the experimental fact that it is very difficult to compare the loudness of tones of different frequency. In such a case the relative magnitudes of  $R$  must be judged.

There then remains the facts of binaural audition and some of the more subtle psychological effects, such as "volume" and "brightness."

In Fig. 16 a diagrammatic sketch of the auditory tract taken from Fig. 13 is shown. The circle represent places where some nerve fibres

stop and others begin. These are scattered along the length of the nerve path. Some nerves have a continuous passage from the cochlea to the auditory brain center, while others may be interrupted three or four times by such synapse. As shown the nerves running from each ear are cross-connected at two points. More nerve fibres coming from the left ear terminate in the right side of the brain and vice versa.

So we see that according to the conceptions given above, when the left ear receives a tone then there should be two space patterns in the brain which are almost alike, one on the left side and one on the right. They are designated by the letter *L*. Similarly, a sound in the right ear produces the two patterns labelled *R*. If the nerve is interrupted at *A*, the left ear will be totally deaf. If, however, it is interrupted at *B* or *C*, only a slight diminution in hearing in either ear will occur. Cases of brain tumor have been found where it was necessary to remove one of the auditory brain centers and also a considerable portion around it and still the patient recovered. Although it left one side of the body paralyzed, hearing tests on such a person indicated that the two ears had practically the same sensitivity as before the operation.

Although some of the nerve fibres from the two ears are placed near together for part of their length and terminate close together in the periphery of the brain, there is little interference between them. This is shown by the experimental fact that a tone introduced in one ear will produce little masking to a sound introduced into the other one. Wegel and Lane showed that a tone to be masked in an opposite ear to that receiving the masking tone needed to be 1,000,000 more intense, that is, at a 60 *db* higher intensity level than if introduced in the same ear as the masking tone.

Also, binaural "objective" beats will not occur until the intensity of tone in one ear is at least 60 *db* above that in the other ear. Objective beats are distinguished from "subjective" beats by the following criteria. They can be readily heard by every observer and can be produced at all audible frequencies while the "subjective" beats can only be heard by about 80% of observers and only then when the frequencies are below 1000 cycles per second. The objective beats are undoubtedly produced by the more intense sound forcing the bones of the head into vibration so that the sound is conducted to the cochlea of the opposite ear. Consequently, such an intense tone produces the same effect as if it were attenuated and introduced directly into the same ear as the other tone.

The patterns from the two ears are laid down close in duplicate so that an accurate comparison can be made. Since the times of maximum stimulation have been shown to be directly correlated to the times of maximum displacement in the medium conducting the sound, the differences of phase which occur at the two ears will manifest themselves as phase differences in the times of maximum stimulation between *L* and *R* on both sides of the brain. In locating the direction of a source of sound both the differences in intensity and also the differences in phase are used. Due to a process of education we have come to associate a definite direction of sound with this difference in the intensity of the stimulation between *L* and *R* and the times between the occurrence of maximum stimulation. It has been seen that the definiteness of this time pattern in the brain decreases with increasing frequency which agrees with the fact that the location of the direction of a source of sound having frequencies higher than 1000 cycles is extremely difficult.

If tones having the same frequency and intensity are applied to the two ears and then the phase changed, it causes a sound image to rotate around the head. Such an effect is related to our experience of locating sounds but it is a new experience since the usual change in intensity is missing, but most persons can make an adjustment for this absence of intensity change and assign a definite direction for the sound image. The changing of phase can be produced by making the tones in the two ears slightly different in frequency. Then the phenomenon known as "subjective" binaural beats is produced. As mentioned above, these beats cannot be heard for tones having high frequencies which is in accordance with the theory outlined.

#### SUMMARY

Summarizing then, the pitch of a tone is determined both by the position of its maximum stimulation on the basilar membrane and also by the time pattern sent to the brain. The former is probably more important for the high tones and the latter for the low tones. The loudness is dependent upon the number of nerve impulses per second reaching the brain and possibly somewhat upon the extent of the stimulated patch. The experience called by psychologists "volume" or "extension" is no doubt identified with the length of the stimulated patch on the basilar membrane. This extension is carried to the brain and forms a portion of excited brain matter of a definite size. It is then this size that determines our sensation of the "volume" of a tone. The low pitched or

complex tones have a large "volume" while the high pitched tones have a small one.

The psychological experience called "brightness" may be identified with the sharpness of the peaks in the vibration form of the basilar membrane as suggested by Dr. Troland. The high tones give the sense of brightness while the low tones the sense of dullness which as seen from Fig. 10 corresponds to the sharpness of the peaks in the vibration form.

The time pattern in the air is converted into a space pattern on the basilar membrane. The nerve endings are excited in such a way that this space pattern is transferred to the brain and produces two similar space patterns in the brain, one on the left and the other on the right side. Enough of the time pattern in the air is sent to each of these stimulated patches to make times of maximum stimulation in each patch detectable. So when listening to a sound with both ears, there are four space patterns in the brain produced, each carrying also some sort of time pattern. It is a recognition of the changes in these patterns that accounts for all the phenomena of audition.