Auditory Patterns*

HARVEY FLETCHER
Bell Telephone Laboratories, New York, New York

DURING the last two decades considerable progress has been made in understanding the hearing processes taking place when we sense a sound. The application of the same instrumentalities that have brought such a wonderful development in the radio and sound pictures to this problem is largely responsible for this progress. Such instrumentalities have made it possible to make accurate measurements which are the basis for understanding any physical process.

To understand this problem then we need to know first how to describe and measure the sound reaching the ears; then we need to know how to describe and measure the sensations of hearing produced by such a sound upon a listener. To do this quantitatively we must also know the degree and kind of hearing ability possessed by the listener. It is with these three phases of the problem that this paper deals.

A pure tone can be defined physically by giving the intensity, the frequency and the phase of vibration at the position where the head of the listener is to be placed. To be precise, this position is taken as the middle of a line connecting the two ears. The measurement of course is made before the head is placed in this listening position. Then the listener faces the source of sound. Any modifications of the sound wave produced by introducing the head to this position are considered as modifications produced by the hearing mechanism. The phase changes are usually unnoticed by the listener, so for most work only the intensity and the frequency are given. A pure tone then is specified by the two coordinates, intensity I and frequency f. In most experimental work on pure tones an endeavor is made to use a free progressive wave or its equivalent, that is, one which is free from reflected waves. Under such circumstances, if I is measured in watts per square centimeter and the corresponding pressure variation in the air wave, expressed in dynes per square centimeter, is p, then I and p are related by the equation[1]

$$I=(p^2/4)\times 10^{-8}. \tag{1}$$

The intensity I is usually determined from this equation after making experimental measurements of p. To express the values for all pure tones in the audible range requires a very large range of values of I and p, so a new term called the intensity level and designated by the letter β has been found useful. It is related to I and p by the equation

$$\beta=10 \log (I/I_0)=20 \log (p/p_0), \tag{2}$$

where the values

$I_0=10^{-16}$ watt per square centimeter,
$p_0=0.0002$ dyne per square centimeter,

have been adopted as international standards. The intensity level is expressed in decibels and varies over a range of only about 130 db when dealing with acoustic problems.

In Fig. 1 is shown a chart on which may be

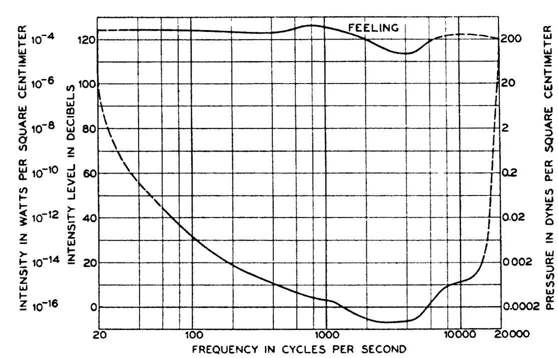

FIG. 1. Auditory area between threshold of feeling and the threshold of hearing.

* From an address given before the joint dinner of the American Physical Society and the American Association of Physics Teachers held in Washington, D. C., on December 28, 1938.

[1] This relation holds exactly for only one temperature and pressure. The variations due to changes in pressure and temperature are small compared to the range in values used in acoustics and for most practical calculations can be neglected.

represented all the pure tones which are audible. It will be noticed that there are three scales for the ordinates and they correspond, respectively, to I, β and p. The values of the frequency f in cycles per second are given by the abscissae. The upper and lower curves enclose the area corresponding to tones which can be sensed as sound. These results are for a young observer having acute hearing. The values of intensity level and pressure are the values obtained in a free air space before the head of the observer is introduced to hear the sound and in a location where the sound waves reflected from walls or other obstacles are negligible. The lower curve gives the faintest sound that such an observer listening with both ears can hear and the upper curve gives the loudest sound that a typical ear can tolerate.

It will be seen from this chart that the pitch range is from 20 cycles per second to 20,000 cycles per second, or a ratio of 1000 to one, and that the intensity range is from 10^{-16} watt per square centimeter to 10^{-4} watt per square centimeter, or a ratio of one thousand billion to one. It will also be seen that the pressure range is from 0.0002 dyne per square centimeter to 200 dynes per square centimeter, or a ratio of one million to one. For a 60-cycle tone, at its maximum tolerable intensity, the amplitude of the air particles is about $\frac{1}{5}$ millimeter. It will be readily seen from these figures that the power necessary to fill a hemisphere 15 meters from a loudspeaker with the maximum sound intensity which can be tolerated by the ear is about 1.5 kilowatts. Of course the electrical power necessary to drive a loudspeaker producing such sounds may be three or four times this value, depending on the efficiency of the loudspeaker. So it is apparent that, although powers involved in sounds are ordinarily very small, still to produce in a large room the loudest sound that the ear can tolerate requires kilowatts of sound power.

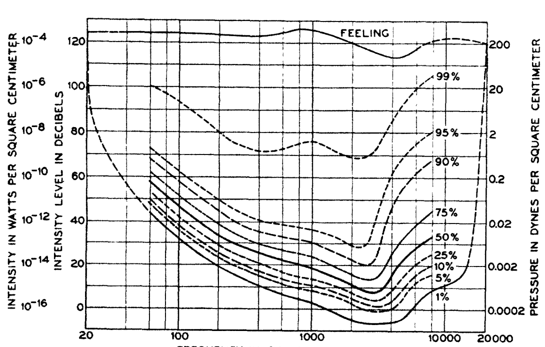

FIG. 2. Auditory areas when hearing is impaired. The curves are indicated by the percent of a typical American group who can hear sounds below the given level.

As stated before, this curve is for an individual having very acute hearing. As a matter of fact about only one in one hundred persons will have as acute hearing as that shown. Recently the Bell Telephone Laboratories cooperated with the U. S. Public Health Service in making a survey of the hearing acuity of a typical American group. The results of this survey have been reported recently by the National Institute of Health in a series of bulletins called the "Hearing Study Series." The results are given in relative intensity levels. Measurements made in our laboratory by Mr. Munson made it possible to reduce these relative levels to absolute intensity levels. From these results the chart shown in Fig. 2 was constructed.

In a general audience, if no noise were present, one percent of the listeners could hear sounds as soft as indicated by the first curve; the next contour line is for 5 percent, and so on for the other curves as indicated. The important curve is the 50 percent curve which is shown by the heavy line. This means that half the people can hear tones at a level as low as indicated by this curve, but the other half must have higher intensities. However, in most auditoriums there is always some background noise. Measurements have shown that in such rooms the threshold values even for acute ears are raised to somewhere near the 50 percent line.

We have thus seen how we can represent a pure tone by a point on the chart such as Fig. 2 by giving the frequency in cycles per second and the intensity level in decibels. Most musical tones have a fundamental and a series of harmonics, so to represent them quantitatively we need to specify the intensity level and frequency for the fundamental and each harmonic. There is a large class of sounds, however, that cannot be specified in such a simple manner. These sounds have components scattered throughout the whole audible range of frequencies. Such sounds are represented physically by giving what is called a spectrogram of the sound. For example, consider a noise like that arising in a busy street. A spectrogram for such a noise is shown in Fig. 3. This spectrogram is determined experimentally as follows. The noise is picked up by a microphone, amplified, and then sent through a filter which permits only a small band of frequencies to pass. The sound intensity of this small band is measured and divided by the width of the band in cycles. This result is then reduced to decibels above 10^{-16} watt per square centimeter to obtain the ordinate in our spectrogram. Instruments are available which will read directly the ordinates in such a spectrogram. The total intensity level is given by the solid bar at the 85-db level.

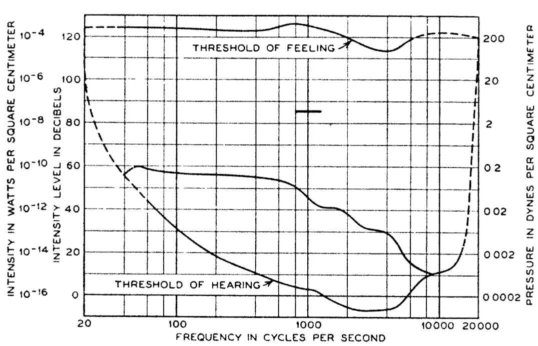

FIG. 3. Spectrogram of street noise.

Similarly, it has been found convenient to give a spectrogram for orchestral music. It is true, of course, that such a spectrogram, to be a true representation of the music, would be continually varying. However, when the full orchestra is playing the variations about the spectrogram curve shown in Fig. 4 are only moderate.

Now we will ask ourselves the question, "What is going on in our ears as we sense such sounds?" To understand more clearly what follows it is necessary to give here a brief description of the hearing mechanism.

The main parts of the hearing mechanism are shown in Fig. 5, the inner ear being shown on a very much enlarged scale. As indicated in this diagram, it is usual to divide the mechanism into three parts—the outer ear, the middle ear and the inner ear. The outer ear consists of the external part, or pinna, and the ear canal which leads into the drum. Beyond the drum is the middle ear which contains three small bones called the hammer, the anvil and the stirrup. The stirrup in turn is attached to a little membrane called the oval window which separates the middle ear from the inner ear. The inner ear is filled with a fluid. As illustrated in the diagram the inner ear is composed of two canals separated by a membrane, all of which is wound into the form of a spiral. The nerve endings are scattered along the membrane separating the two canals. Very tiny nerve fibers coming from the main auditory nerve are fanned out so as to reach each of these nerve endings. The arrangement of the nerve fibers is similar to the arrangement of wires in a telephone cable which are fanned out and connected to the jacks in a telephone switchboard. In such a switchboard, used to its full capacity, there are 10,000 jacks and to each one three wires are connected called the tip, ring and sleeve, making a total of 30,000 wire terminals.

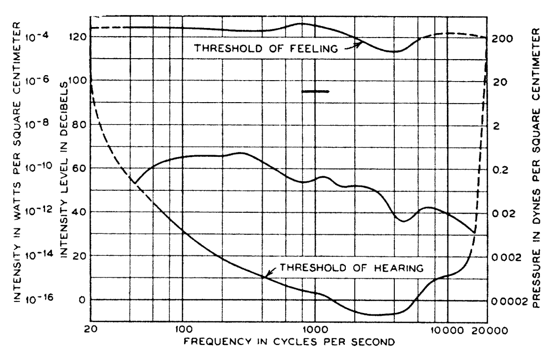

FIG. 4. Spectrogram of orchestral music.

There are approximately this number of nerve terminals along the basilar membrane and nerve fibers leading from them to form a bundle called the auditory nerve. In telephone engineering circles it has been considered a real achievement to have developed a design which permits 30,000 wire terminals to be within the reach of a single operator. The area of such a switchboard panel is about 10,000 square centimeters. Nature has accomplished a similar thing in the hearing mechanizm occupying an area of only $\frac{1}{10}$ of a square centimeter, or only one hundred thousandths of the space in a telephone switchboard. The nerve terminals occupy a space only about $\frac{1}{3}$ millimeter wide and 30 millimeters long. Each one of these tiny nerve endings acts like a telephone transmitter. When they are agitated by the sound waves coming into the inner ear they transmit an electrical current along the nerve fiber. It is these electrical currents which go to the brain and cause the sensation of hearing. The sound wave passes down the ear canal, strikes the drum of the ear at *t*, vibrates the drum, which in turn vibrates the anvil, which in turn vibrates the stirrup. The stirrup then vibrates the oval window which transmits the vibrations to the fluid of the inner ear. This fluid transmits the vibration to the nerve terminals, which in turn converts the vibration into an electrical disturbance which is sent to the brain.

In the following discussion it is not necessary to know the details of this mechanism of hearing. It is important, however, to know that the nerve endings are scattered along a space which is one hundred times as long as it is wide. For this reason we can designate the position of a small group of nerves by a single coordinate. Also, since it is known that this long and narrow membrane containing the nerve endings is wound into a spiral, I have chosen to represent the position of the nerves on such a spiral diagram. A typical form of this is shown in Fig. 6. The distance along this spiral from one end to the other in a typical human cochlea is about 3 centimeters. The numbers along the outside run from zero to 100. In talking about the positions in this cochlear diagram I will refer to these small numbers between zero and 100 as positions of nerve

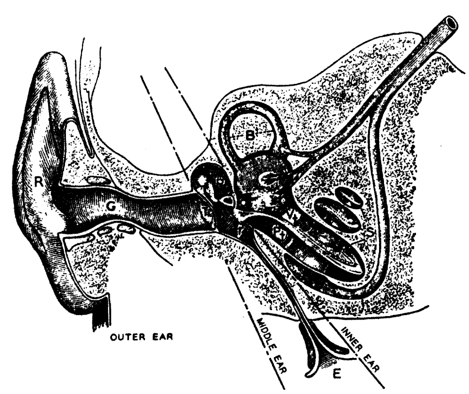

FIG. 5. Semi-diagrammatic section through the right ear (Czermak). *G*—external auditory meatus; *T*—membrana tympani; *P*—tympanic cavity; *o*—fenestra ovalis; *r*—fenestra rotunda; *B*—semi-circular canal; *S*—cochlea; *Vt*—scala vestibuli; *Pt*—scala tympani; *E*—Eustachian tube; *R*—pinna.

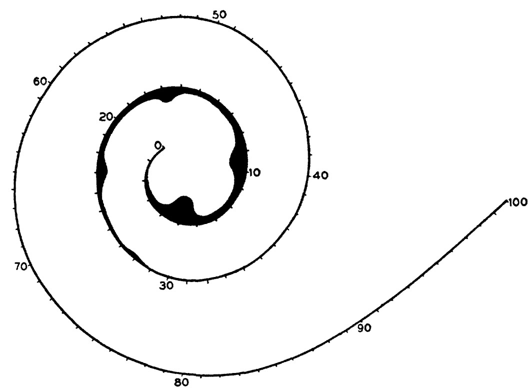

FIG. 6. Typical auditory pattern.

endings, or more briefly as nerve positions. Also for convenience I will speak of a group of nerve endings comprising one percent of the total as a patch of nerves. So the position of each patch is designated by one of the numbers between 0 and 100.

On the chart of Fig. 6 we see a diagram which I will call an auditory pattern. On this spiral diagram representing the cochlea, the height of the dark portion is drawn proportional to the response of the nerves at that position. The pattern shown is that produced in a typical normal ear when a 200-cycle tone is heard at an intensity level of 90 db. The response is measured as the amount of loudness going from each nerve patch at the positions indicated. There is good evidence, although not conclusive, that this loudness is proportional to the number of nerve impulses being sent to the brain per second by each patch of nerves. For example, you see that most of the loudness comes from the region around position 5. However, there is considerable part of the loudness coming from other parts corresponding to the subjective harmonics; for example, at positions $10\frac{1}{2}$, 17, 23 and 28. These subjective harmonics are introduced by the mechanism of hearing. The total loudness corresponds to the total area of the darkened part and this corresponds to the total number of nerve impulses reaching the brain along the auditory nerve.

Later I will discuss methods of deducing such auditory patterns from experimental measurements, but before doing this I wish to present the auditory patterns corresponding to the pure

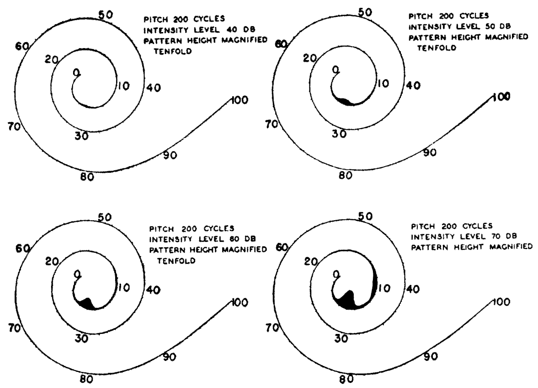

FIG. 7. Auditory patterns for a 200-cycle tone at various intensity levels.

tones. In Figs. 7 and 8 are shown such patterns for a 200-cycle tone at the various intensity levels shown. For intensity levels below 70 decibels the response is confined almost entirely to the positions between 4 and 5. As the intensity level goes above this value it will be seen that more and more subjective harmonics begin to appear until at an intensity level of 110 decibels there is response of the nerves all along the first turn and a half of the cochlea.

In Figs. 9, 10 and 11 are shown the auditory patterns for tones having an intensity level of 90 db but varying in frequency throughout the audible range. It will be seen that the tones of low pitch stimulate the nerves at positions corresponding to low numbers, that is, near the end of the inner spiral called the helicotrema. As the frequency of the tone increases the position of maximum response gradually shifts to positions corresponding to higher and higher numbers. At the same time the position of response due to the harmonics continues to shift as the pitch increases.

How do we know that these auditory patterns correspond to what is taking place in our ears? To determine these patterns we need to know the amount of response as represented by the breadth of the dark portion in the auditory pattern and the position along the spiral where the response is occurring, that is, the nerve position. The fundamental data from which such patterns are drawn are obtained from the masking effect of such sounds. For example, in Fig. 12 are shown such data on the masking produced by pure tones. At the top of this figure is shown the masking produced by a 200-cycle tone. The ordinates give the masking or the number of decibels that the intensity level of a tone having the frequency shown by the abscissa must be raised above its quiet threshold in order to be heard in the presence of this 200-cycle sound. The peaks[2] at 400, 600, 800, 1000 are due to the subjective harmonics. For example, consider the point at 700. The masking here is seen to be 22, which means that a tone of 700 cycles must be raised 22 db above its quiet threshold to be heard in the presence of a 200-cycle tone. Such a curve showing the masking at each frequency is called a noise audiogram.

In the lower part of this figure the masking for several tones of different frequency is given. Only that part near the fundamental is shown. For the sake of clearness the peaks due to the harmonics are omitted. I wish to call attention particularly

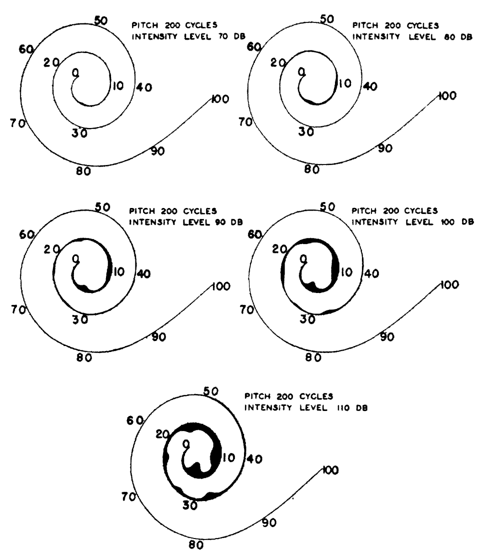

FIG. 8. Auditory patterns for a 200-cycle tone at various intensity levels.

[2] There is an uncertainty of 4 or 5 db in determining these peak values, which is not due to observational error but due to interpretation. When testing the masking at frequencies near the fundamental or any of its subjective harmonics, then beats occur. This new phenomenon makes it much easier to detect the masked tone. The points given are for most distinct beats rather than where the beats just become audible. The points should be somewhere between the best beat and minimum audible beat point. To enter into the discussion here of the best point to take would tend to confuse the main argument of the paper. Changes due to such uncertainty would not change the general shape of the auditory patterns but only alter somewhat the sharpness of the peak.

to the fact that the breadth of these masking curves increases rapidly as the frequency of the tone doing the masking goes above 1000 cycles. Of course this must be related to how the ear mechanism analyzes the sound. The bars at the top of these curves are the minimum frequency band widths of thermal noise for masking these tones and will be discussed later.

In Fig. 13 is shown the masking effect of a noise whose importance will become evident as

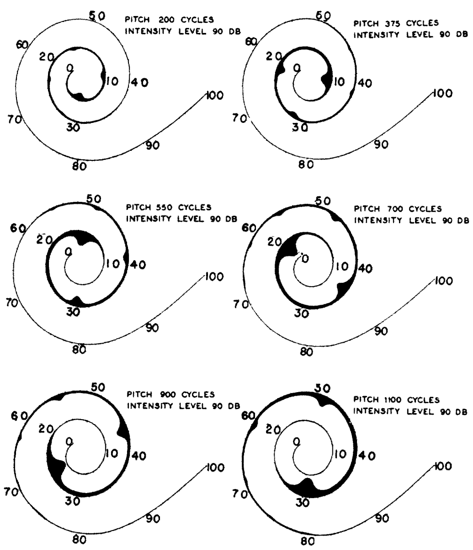

FIG. 9. Auditory patterns at 90 db for various frequencies.

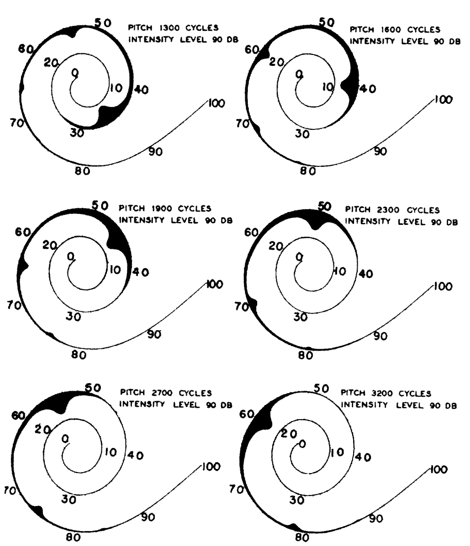

FIG. 10. Auditory patterns at 90 db for various frequencies.

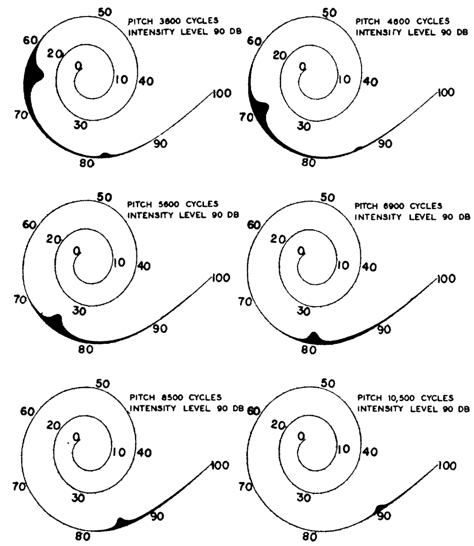

FIG. 11. Auditory patterns at 90 db for various frequencies.

the discussion proceeds. The curve at the bottom of the crosshatched area is the threshold of hearing curve for a quiet place and is taken from Fig. 1. The curve at the top of the hatched area is the spectrogram for this noise. If I_f is the intensity per cycle of the noise for the frequency region f, then the ordinate for this spectrogram curve is given by

$$y = 10 \log (I_f/I_0) \tag{3}$$

and the total sound intensity I of the noise is given by

$$I = \int_0^\infty I_f df = I_0 \int_{20}^{20,000} 10^{(y/10)} df \tag{4}$$

and the intensity level β by

$$\beta = 10 \log \frac{I}{I_0} = \left[10 \log \int_{20}^{20,000} 10^{(y/10)} df \right]. \tag{5}$$

This value calculated from the spectrogram is shown by the solid bar placed across the 1000-cycle ordinate.

The solid dots show the intensity levels of pure tones which are just masked by this noise. The number of decibels these dots are above the threshold curve gives the masking. It will be seen that this is the same for each dot and is equal to 50 db. In other words, the masking for

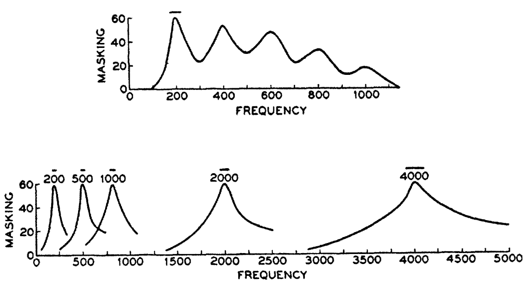

FIG. 12. Noise audiograms of pure tones.

this particular kind of noise is the same for all pure tones. The masking audiogram will then be very simple, that is, a straight line at 50 db for all frequencies.

If we assume that constant masking indicates constant stimulation along the different patches of nerves, then it is evident that the auditory pattern corresponding to this noise giving constant masking for all frequencies is similar to that shown in Fig. 14. It is because such a noise produces a simple pattern that it becomes very important in our discussion.

It is from masking curves like those shown in Fig. 12 and the constancy of the masking by auditory nerves that the auditory patterns can be drawn. To do this we must know how the masking in decibels, that is the ordinate in these audiograms, is related to loudness per patch of nerves, and also how the frequency, that is the abscissae of these audiograms, is related to the nerve position.

First let us consider the data which enable us to determine the relation between the frequency

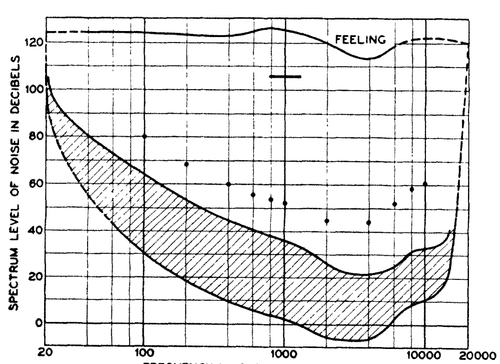

FIG. 13. Masking effect of a noise.

of the sound and the position of maximum stimulation in the cochlea. The nerve position will be designated by the coordinate x (see Fig. 6) going from 0 to 100, and the frequency of the sound will be designated by f. Our problem then is to find the relation between x and f. There are three experiments dealing with quite different phases of audition which give us data from which this relation can be calculated. The first set of data is obtained from experiments on the minimum perceptible differences in pitch. In these experiments the frequency of the tone is varied until the observer perceives a change in pitch. The technique of doing such experiments is

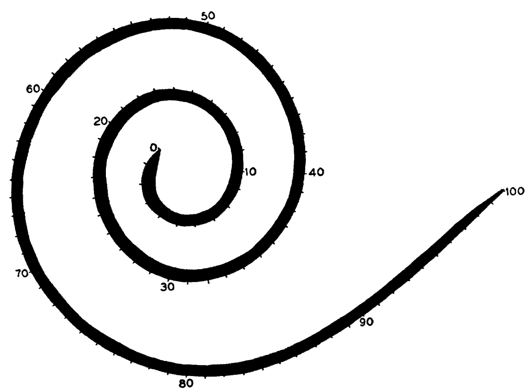

FIG. 14. Auditory pattern stimulation uniform at all frequencies.

described elsewhere.[3] The results obtained by Shower and Biddulph are given in Fig. 15. The ordinate gives the value of the cycles change for a tone to be just audible and the abscissa gives the frequency of the tone. As you will notice, tones whose frequencies are below 1000 cycles require about the same number of cycles for a minimum perceptible change, namely, between 2 and 3 cycles change, while for higher frequencies the fractional rather than the actual increase of frequency is constant and equal to about 0.3 percent. From these data we can derive the desired relation[4] between x and f as follows. It was seen that the position of maximum stimulation shifted as the frequency changed. It seems reasonable to assume that the amount of this shift for a minimum perceptible change in fre-

[3] Shower and Biddulph, J. Acous. Soc. 3, 275 (1931).
[4] This relation was first derived by Wegel and Lane of the Bell Telephone Laboratories in 1924.

quency is the same for all positions. Therefore if Δx is this shift, then

$$\Delta x = (dx/df)\Delta f = A \tag{6}$$

where A is a constant for all frequencies. With the scale we have chosen it is the percent of the total nerve endings passed over to make the change noticeable. Then

$$x = A\int_0^f \frac{1}{\Delta f}df. \tag{7}$$

The condition $x=100$ when $f=\infty$ determines the constant A, or

$$x = 100\frac{\int_0^f \frac{1}{\Delta f}df}{\int_0^\infty \frac{1}{\Delta f}df}. \tag{8}$$

From the curve of Fig. 15 the constant A is determined as 0.08 percent corresponding to about 20 nerve endings.

A second set of data for obtaining this relation is obtained by finding the ratio of the intensity per cycle of the noise to the intensity of the pure tone which is just masked by the noise. If we return again to Fig. 13 it will be noticed that in the region below 1000 cycles the distances between the dots and the spectrogram curve for the noise are about the same and equal to 15 db. In

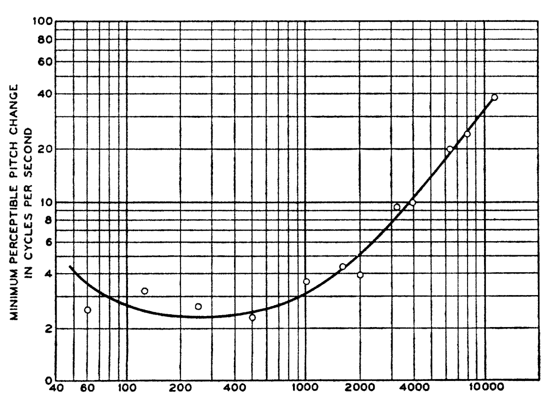

FIG. 15. Minimum perceptible pitch change.

regions above 1000 cycles, however, the distance becomes greater and greater until at 8000 cycles it is 28 db. Now why is this? It is because the frequency positions in the ear are crowded together in this higher pitch region. A wider band of frequencies is going to each patch of nerves corresponding to these higher frequencies and consequently produces a relatively greater stimulation, and consequently greater masking. These ideas can be formulated into a mathematical equation as follows:

Let

ΔI = sound intensity of noise in band between f and $f+\Delta f$;

$I_f = (\Delta I/\Delta f)$ = intensity per cycle of noise at frequency f;

ΔJ = agitating power due to noise on cochlear nerve endings between x and $x+\Delta x$;

$J_x = (\Delta J/\Delta x)$ = agitating power due to noise upon one percent of nerves at position x. It may be called agitation density;

I_m = sound intensity of tone of frequency f which is just perceptible in presence of noise;

J_m = total agitating power on nerve endings at position x by this tone which is masked.

Then if the transmission system is linear $(\Delta I/I_m) = (\Delta J/J_m)$ or, substituting values for ΔI and ΔJ, we have

$$\Delta x = (I_f/I_m)(J_m/J_x)\Delta f. \tag{9}$$

We will now make the following assumption[5]

$$J_m/J_x = C, \tag{10}$$

where C is a constant independent of x and determined by condition $x=0$ when $f=\infty$. Therefore

$$x = C\int_0^f \frac{I_f}{I_m}df \quad \text{and} \quad C = \frac{100}{\int_0^\infty \frac{I_f}{I_m}df}. \tag{11}$$

If we compare the relation given by Eq. (11) to that given by Eq. (8) we must conclude that, since these express the same relation between x and f, the two integrands must be proportional, or

$$\Delta f = k(I_m/I_f), \tag{12}$$

where k is a constant which has been found by experiment to be equal to 20 when dealing with thermal noise.

[5] For a more complete derivation of these relations see Proc. Nat. Acad. Sci. 24, 265–274 (1938).

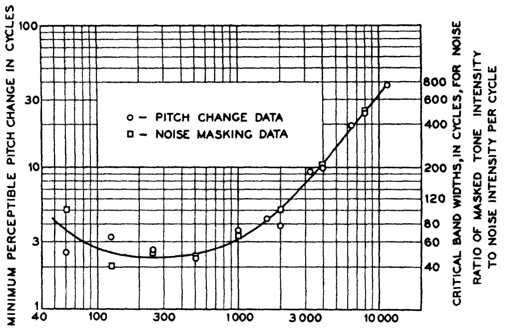

FIG. 16. Plot showing that the minimum perceptible pitch change is proportional to the ratio of the intensity of a tone having the same frequency to the intensity per cycle of the noise which just masks this tone.

In other words, this analysis predicts that the minimum perceptible change in cycles for a tone of any frequency is equal to a constant times the ratio of the intensity of a tone having the same frequency to the intensity per cycle of the noise which just masks this tone. These two sets of experimental data are plotted in Fig. 16 and it will be seen that they do determine the same curve. The masking data were taken from the average of a large number of observations similar to those shown in Fig. 13. If K is the difference in decibels between the dots and the spectrogram curve in this figure, then K is related to I_f/I_m as follows:

$$K = 10 \log (I_m/I_0) - 10 \log (I_f/I_0) \quad (13)$$

or

$$I_m/I_f = 10^{(K/10)}. \quad (14)$$

In obtaining these data the question which led to the third type of experiment arose. It is: how does the ratio I_m/I_f vary as the width of the noise band in the neighborhood of the frequency of the masked tone becomes smaller and smaller. It would be expected that when this width became smaller than a certain limiting value, then all of the acoustical intensity could be considered as acting upon the same nerve endings as those stimulated by the tone being masked. Consequently, for band widths smaller than this limiting value, this ratio I_m/I_f becomes smaller and smaller as the band width becomes smaller and smaller. In fact it would be expected that for these smaller band widths the ratio of the intensity of the masked tone to the total intensity of the noise in the small band would be a constant independent of both the band width and the frequency region in which it is placed. Consequently, for these small band widths the ratio I_m/I_f will be proportional to the width of the band. For all values of band widths larger than this critical one this ratio will remain constant.

Experimental tests were made to verify these relations. I will not take the time to describe the apparatus and methods[6] for making this kind of experiment, but will be content however to show the final results, which are given in Fig. 17.

The abscissae give the width of the noise band and the ordinates the ratio of the intensity of the masked tone to the intensity per cycle of the noise. The variable parameter which identifies the different curves is the frequency of the masked tone, which is also the frequency region of the band of noise. It is evident from the results for the 30-cycle band of noise that this ratio is the same for all frequencies and is approximately equal to 30. This result makes the relations very simple. For these smaller bands and for this type of statistical noise, the intensity of the masked tone must be adjusted to be equal to the average intensity of the noise in the band in order for it to be perceived. It will be seen that as the band widths increase, this ratio also increases until the critical band widths are reached, which are indicated by the intersection of the horizontal lines with the 45-degree line.

For example, consider the case for the 2000-cycle tone being masked by variously sized bands

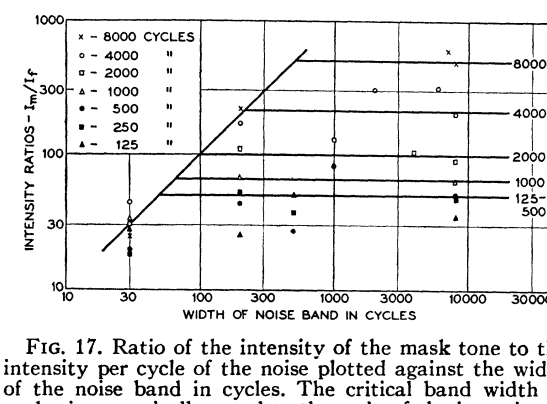

FIG. 17. Ratio of the intensity of the mask tone to the intensity per cycle of the noise plotted against the width of the noise band in cycles. The critical band width in cycles is numerically equal to the ratio of the intensity of the tone masked to the intensity per cycle of the noise producing masking and always corresponds to ½ mm of length on the basilar membrane.

[6] This work will be presented in a future paper.

of noise, indicated by the squares on Fig. 17. It will be seen that the ratio increased until the band width of 100 cycles is reached. Increasing the width beyond this value does not affect the ratio. In other words, increasing the band width[7] beyond 100 cycles does not change the masking at all for hearing the 2000-cycle tone. The horizontal lines were determined from measurements on wide bands of thermal noise and are the same values as were given in Fig. 16. Since their intersection of the 45-degree line determines the critical values of band widths, this very important relation follows:

For this type of noise the critical band width in cycles is numerically equal to the ratio of the intensity of the tone masked to the average intensity per cycle of the noise producing the masking. Regardless of where the band is located, we will see later that these critical widths always correspond to a single element of length on the basilar membrane, namely $\frac{1}{2}$ mm. So we can return again to Fig. 16. The single curve shown can now be given three different interpretations: first, the ordinates can be interpreted as giving the cycles change necessary to produce a perceptible change of pitch; second, when the ordinates are multiplied by 20 they will then give the critical band widths; third, when the ordinates are multiplied by 20 they will give the ratio of the intensity of the masked tone to the intensity per cycle at the frequency of the masked tone of a statistical noise doing the masking. For these reasons we have considerable confidence in the validity of the shape of this curve.

By means of this curve then and either of the formulae developed above, the relationship between x and f can be obtained and is shown in Fig. 18A on a rectangular plot and in Fig. 18B on a spiral diagram. This relation enables us to find the position of maximum stimulation produced by sounds whose power is confined in the region near the frequency f.

Next we will inquire as to what kind of data is used to determine the height of these stimulation patterns. It will be remembered that the height of these auditory patterns is proportional to the loudness being sent to the brain by 1 percent of

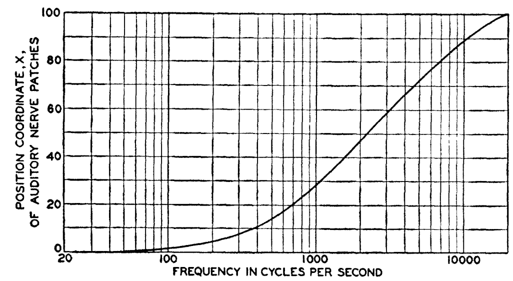

FIG. 18A. Position coordinate X of auditory nerve patches for various frequencies.

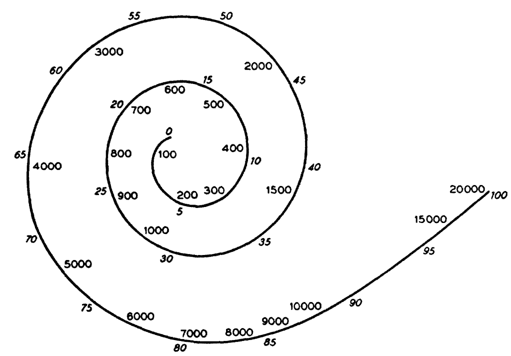

FIG. 18B. Position in the cochlea for maximum response to pure tones.

the nerves at each position. If this is true, then of course the area[8] under the auditory pattern curve will give the total loudness. In order for this to have any meaning we must have a quantitative scale of loudness.

For purposes of making comparisons of loudness of various sounds, a reference sound has been adopted by the American Standards Association as a reference standard. This same standard has recently been recommended for adoption to the International Standards Association by an international acoustical conference. This reference sound is a free progressive wave having a frequency of 1000 cycles per second to be received by the listener facing the source of this tone. Its intensity could be defined by giving either I, p, or β, but it has been agreed internationally that it shall be defined by giving the intensity level in decibels. The loudness level of the reference tone expressed in phons is the intensity level of this tone expressed in decibels. The loudness level of any other sound is determined by comparing

[7] It is realized that these critical band widths are not accurately determined from these data. They may be wrong by a factor of 2, but that the bands are the right order there is no question.

[8] For this to be strictly true the patterns should be plotted in rectangular coordinates.

it to the reference tone, adjusting the latter until the two sound equally loud. Then the loudness level of the reference tone becomes also the loudness level of the sound being measured and is expressed in phons.

Loudness is the magnitude of the auditory sensation produced by the sound. A scale for representing loudness has been found such that units on this scale agree with common experience in the estimates made of the sensation magnitudes. This scale is definitely related to the loudness level scale. Units on the scale are called loudness units, abbreviated LU.

A true loudness scale must be one such that when the number of units on this scale is doubled, the magnitude of the sensation as experienced by typical observers will be doubled, or when trebled the loudness sensation is trebled, etc. The existence of such a scale is dependent upon the supposition that consistent judgment tests of loudness can be obtained. Experimental tests indicate that such judgments are sufficiently consistent to warrant choosing such a scale. To find the relation between the loudness level and the loudness on the proposed new scale, various types of experiments were performed by investigators in our laboratory and elsewhere.

The general scheme of these experiments is to find experimentally pairs of values of loudness level L_1 and L_2 phons, which correspond to pairs of values of the loudness N_1 and N_2 LU. The first type of experiment is concerned with a comparison of the loudness experienced by a typical observer when listening first with one ear and then with both ears. Since we are dealing with the typical observer, the two ears will be alike in sensitivity. Consequently, if loudness is proportional to the total nerve energy sent to the brain, then it follows that the loudness experienced when using two ears will be double that when using one ear. To obtain such data the sound is first produced at any convenient intensity and listened to with only one ear, and its loudness level measured. Let us call this value L_1. Without changing the intensity of the sound the observer listens with both ears and the loudness level is measured. Let us call its value L_2. If N_1 is the number of loudness units corresponding to listening with one ear and N_2 the number corresponding to listening with both ears, then it is evident that N_2 must be just double N_1. We will again listen with one ear, but increase the intensity of the sound until a loudness level of L_2 is obtained. If now again the intensity of the sound is held constant and we listen with both ears, we will find a loudness level L_4. This must correspond to a loudness N_4 which is four times N_1. And so we might continue throughout the entire intensity range. If we now choose N_1 arbitrarily equal to some convenient number, we will have a set of corresponding values of loudness units and loudness levels from which we can draw a curve giving the relation which we are seeking.

The experiments as outlined could not be carried out because our typical observer was hypothetical. To obtain results corresponding to that which would be obtained by such a typical observer, one must deal with a large group of observers, taking the average value as that corresponding to a typical observer. For this reason, pairs of loudness levels L_1 and L_2 were obtained starting with L_1 at various values throughout the audible range. This experiment was tried with pure tones having frequencies of 100, 125, 500, 1000, 2000 and 4000 cycles per second, respectively. The indicated points in Fig. 19 show the observed values obtained. A solid curve is drawn to give the best fit of the observed points. From this curve the proper values of L_1, L_2, L_4, L_8, etc., can be obtained to build the scale in the manner described above.

In the second type of experiment the following procedure is used. A 1000-cycle reference tone is balanced against a pure tone having a frequency far removed from 1000 until they are equally loud. Then the two tones are combined to form a complex tone having two components. Let L_1 be the loudness level of each component alone, and L_2 the loudness level of the combination. Since the components were adjusted to produce equal loudness, it is assumed that each is producing the same nerve stimulation and that when they act together they will produce twice the loudness. This hypothesis is reasonable provided that in the mechanism of hearing the two sets of nerves do not interfere with each other. The auditory patterns given for pure tones show that such interference is negligibly small when their frequencies are widely separated. Consequently,

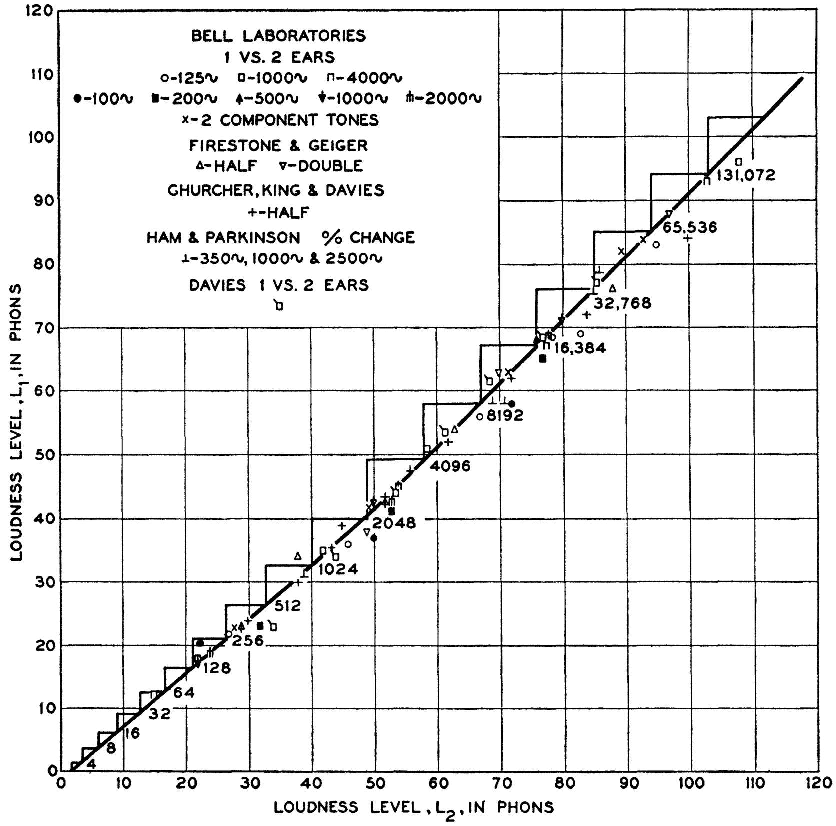

FIG. 19. Loudness determinations; various experimenters.

balancing such a two-component tone against the reference tone should give pairs of values of L_1 and L_2 which can be used in the same way as indicated by the one-*vs.*-two-ear method described above, that is, L_2 corresponds to a loudness number which is just twice that corresponding to L_1. Such observations are shown in Fig. 19 by the points indicated x.

In the third type of experiment a typical observer is asked to listen to a sound, first at a loudness level L_1 and then to the same sound when its loudness level has been raised to a value L_2 such that to him the loudness seems to have doubled. The experiment is repeated, starting with another initial value of L_1, and thus a series of pairs of values of L_1 and L_2 are obtained which have the same relation to our loudness scale as those obtained by the first two methods. A much greater amount of data is required by this method because the probable error of a single judgment is very large. However, this method is related more directly to the scale we are seeking than the two preceding ones. Four different investigators[9] in different parts of this country and England have taken a very large number of observations by this method, not only on the 1000-cycle reference tone but on many types of tones. The data taken by Laird have been omitted since they were so far out of line from those obtained by the other three groups of investigators, namely, Ham and Parkinson, Firestone and Geiger, and Churcher-King-Davies. Their observations are given in Fig. 19 by the points indicated.

It will be seen that the observed points obtained by these diverse methods fit consistently on one curve, particularly well in the range of loudness usually experienced in everyday life;

[9] In these experiments the loudness level was not measured, but only the number of db above the threshold. For the 1000-cycle tone this was taken as the loudness level L. For other pure tones the loudness level was computed from our data on the loudness level of pure tones.

namely, from $L_2=40$ to $L_2=80$ phons. It was sufficiently consistent to justify choosing the resulting loudness scale as a true loudness scale.

From this curve in Fig. 19, pairs of values of L_1 and L_2 were taken to construct the curve of Fig. 20 in the following way. It will be seen that a staircase has been erected upon the curve with the corners of the steps lying on a 45-degree straight line. The value of N for $L=0$ was taken as unity. If, then, the sound is adjusted so that the threshold is reached for either ear listening alone, then when listening with both ears, the loudness level is between 2 and 3 phons and the loudness is doubled, or 2. This corresponds to the bottom step. Now if the sound is adjusted so that for each ear alone the loudness level is 2 phons, then if both ears are used the loudness is raised to 4 LU. This corresponds to the second step. So it is seen that the width of the steps shows the change in the loudness level necessary to double the loudness. The values of the loudness N indicated on each step can now be plotted against the corresponding L to give the desired relation between L and N. This is shown by the curve of Fig. 20.

This same scale can be developed from data obtained by balancing a three-component tone against the reference tone or by judging when a threefold increase in loudness has been produced. Or we could use any other number besides 2 or 3 and get the same scale if it is a real loudness scale. An extreme case would be to use a ten-component tone and use judgment tests

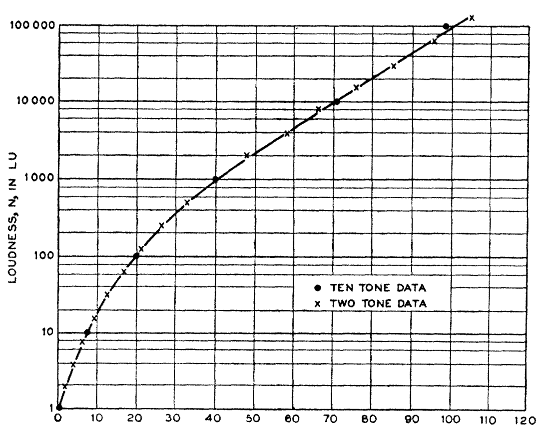

FIG. 20. Loudness scale.

for one-tenth and ten times as loud. Tests of this kind have been made which I shall now describe.

A complex tone[10] was generated having ten harmonics of a fundamental frequency of 530 c.p.s. The components were each adjusted so that when listened to separately they had the same loudness level L_1 phons. They were then combined into a complex tone, and it was found that the loudness level was raised to a value L_{10} phons. So in this way pairs of values of loudness levels L_1 and L_{10} were obtained starting with L_1 at different intensities of sound. If the patches of nerve stimulation corresponding to each component do not overlap, then the complex tone should be ten times as loud as each component. Such pairs of intensities were obtained with a group of observers, and are shown by the points in Fig. 21. Ham and Parkinson, and Firestone and Geiger made judgment tests to find a pair of loudness levels L_1 and L_{10} such that the tone for L_1 was one-tenth the loudness of that for L_{10}. Their results are shown by the points indicated. The points are somewhat scattered but I think it is remarkable that the two sets of points determine the single curve within the observational error, for such judgment tests are extremely difficult to make. The three points for the ten-tone data at the higher levels are definitely off the curve as drawn and should be, for at these high levels the auditory patterns for the different components overlap. The judgment tests and the ten-tone balance tests were all made in different laboratories and before any loudness scale of the type discussed here was constructed.

Now the curve of Fig. 21 can be used to construct a loudness scale. In this case the width of the steps of the staircase corresponds to the phons increase necessary to make the sound ten times as loud. The corresponding values of N and L can be used to determine the curve of Fig. 20. The solid curve was drawn so that points on it would yield the same loudness scale as shown in Fig. 20. It is seen that such a curve does indeed fit the observed data. The following equation

[10] The description of this generator is given in the paper by Fletcher and Munson, J. Acous. Soc. Am. 5, 82–108 (1933).

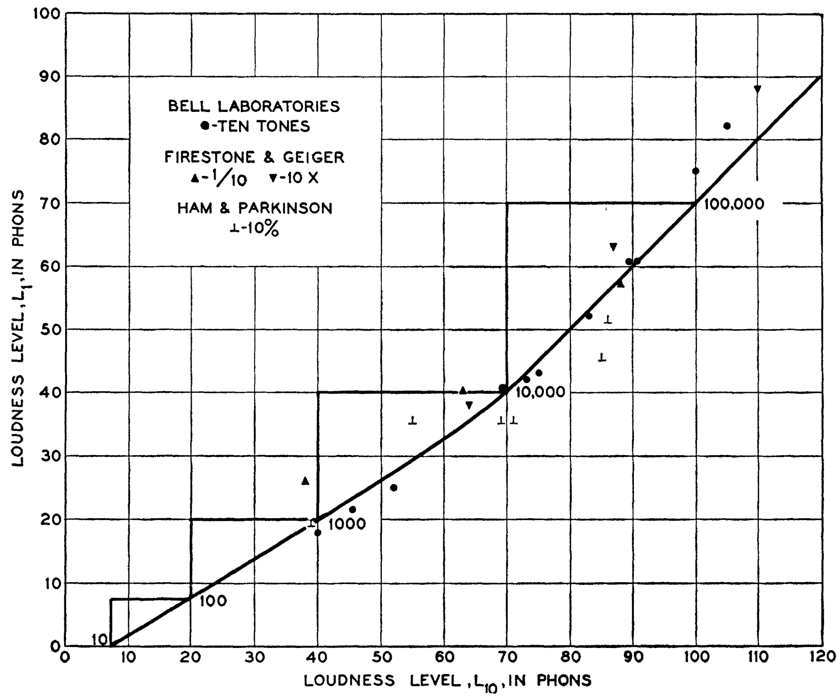

FIG. 21. Loudness determinations; various experimenters.

$$N=\frac{10^{5/3}I}{(I+10^{5/2})^{2/3}}=\frac{10^{5/3}10^{L/10}}{(10^{L/10}+10^{5/2})^{2/3}} \quad (15)$$

proposed by Knauss represents approximately the relation shown in the curve of Fig. 20. This reduces to

$$N=I=10^{L/10} \quad (16)$$

for values of L below 20; that is, at very low levels the loudness numbers are equal to the intensity, where I is expressed in terms of the reference intensity I_0.

$$N=10^{5/3}(I)^{1/3}=10^{5/3}10^{L/30}=46\times 10^{L/30} \quad (17)$$

for values of L above 40. This last relation fits the curve more accurately, the other two only approximately. The last relation states, throughout most of the range ordinarily encountered in practice the loudness numbers are proportional to the cube root of the intensity.

It is important to remember that the values of the intensity I in the above equation refer only to the 1000-cycle reference tone but the values of L are for any type of steady sound. The curve of Fig. 20 has been taken as defining the relation between the loudness N expressed in LU's and the loudness level L expressed in phons. The loudness of any other tone, either simple or complex, can be determined by comparing it to this reference tone and adjusting the intensity of the latter until the two sound equally loud. The loudness of the reference tone thus adjusted is then the loudness of the unknown sound being measured. Stevens[11] has proposed using the word "sone" for a loudness unit 1000 times that called the LU in this paper. Consequently the LU would be the same as a millisone. Loudness levels below 40 db are very faint sounds. For example, the tick of a high quality watch held close to the ear has a loudness level between 40 and 50 phons. The curve in Fig. 22 shows the relationship between the loudness in sones and the intensity level of the reference tone expressed in decibels, both plotted on a linear scale. If one were listening to a 1000-cycle tone and increased the intensity level uniformly by turning a dial of an attenuator

[11] *Hearing, Its Psychology and Physiology*, Stevens and Davis, p. 119.

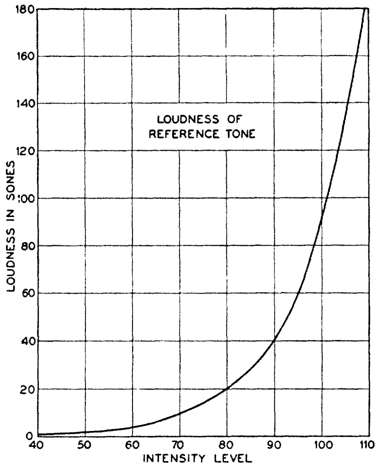

FIG. 22. Loudness of reference tones in sones.

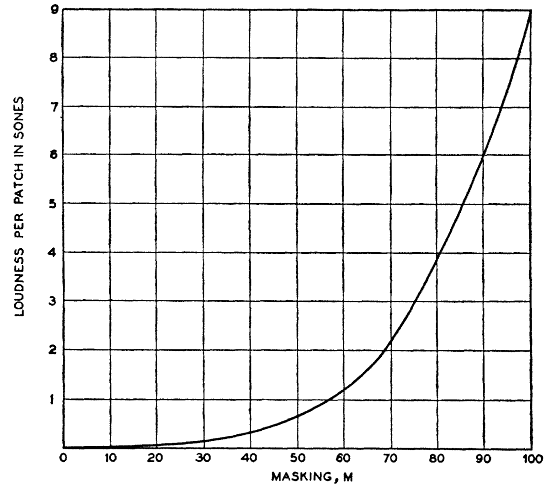

FIG. 23. Relation between the masking, M, and the loudness per patch in sones.

calibrated in decibels, then the loudness increase as sensed by the ear would correspond to the increase in height of this curve as sensed by the eye. One can easily perform this experiment and satisfy himself that this is indeed true.

We now want to use these loudness relations to determine the loudness of our important type of noise, that corresponding to the spectrogram shown in Fig. 13. It will be remembered that such a noise produces a uniform response of all the nerve patches. Consequently, measurements of loudness and masking were made of this noise. The intensity of the noise was then changed without changing the shape of the spectrogram. It was found that within the observational error the masking was again the same for each frequency. Another pair of values of loudness and masking were measured. This was continued until such pairs of values were obtained throughout the entire audible range. Since all the nerve patches are responding equally, if the total loudness is divided by 100 we obtain the loudness per patch. The results of these measurements are given in Fig. 23. This curve then gives us the necessary relation between the masking M and the loudness per patch. In other words, it enables us to know how high to draw the auditory pattern at each position corresponding to the frequency of the tone being masked by the amount M.

Summarizing the procedure then, for any sound we first determine experimentally the masking audiogram. This gives the masking M for each frequency. The frequency scale is transferred to the nerve position scale by the relation of Fig. 18A, and the masking scale to a loudness per patch scale by the relation of Fig. 23. The results are plotted on the spiral diagram to give the auditory patterns corresponding to the noise measured. For example, in Fig. 24 is shown the masking audiogram for a steamboat whistle and in Fig. 25 is the corresponding auditory pattern. In Figs. 26 and 27 are similar diagrams for street noise. Notice that all of the nerve patches are being continually stimulated when one is in such a noise. In Fig. 28 are given the auditory patterns corresponding to the notes of a bugle playing taps. Notice that the harmonics in the

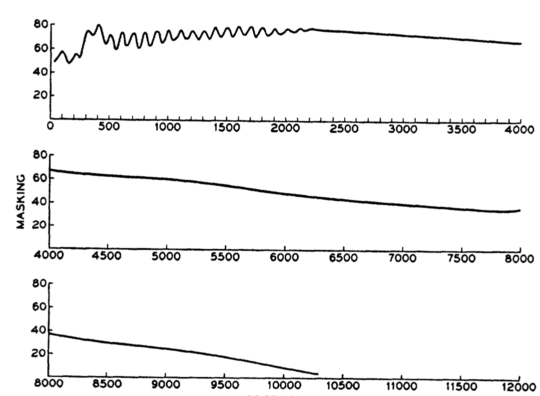

FIG. 24. Masking audiogram for steamboat whistle.

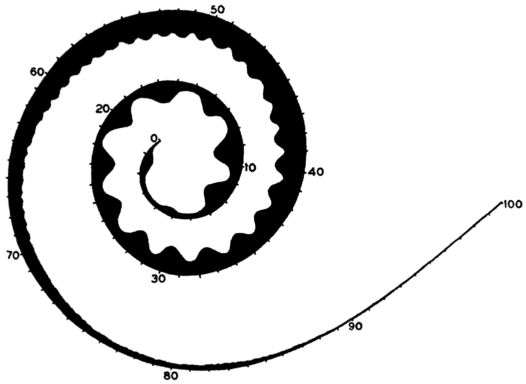

FIG. 25. Auditory pattern of steamboat whistle.

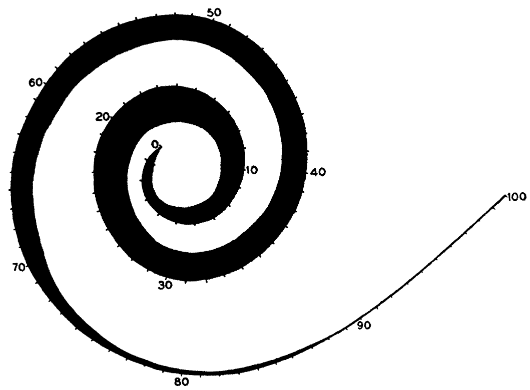

FIG. 27. Auditory pattern of street noise.

positions between 40 and 60 carry most of the loudness. For example, in the first note the 6th, 7th and 8th harmonics each carry about three times as much loudness as the fundamental. For each note the fundamental carries only a small fraction of the total loudness. This is due both to the mechanism of the cornet and also to the mechanism of hearing.

So it is the interpretation of these continually changing auditory patterns in the ear which brings to us all of our information about the acoustic world about us. These patterns have been those experienced by a typical listener who has normal hearing. The question now arises as to what patterns are experienced by a deafened listener. There are two kinds of deafness, which are usually designated "nerve deafness" and "obstructive deafness." In the former the nerve transmitters are defective, while in the latter the mechanical transmitting mechanism from the outside air to the nerves of the inner ear is defective. Of course, both types of deafness can occur in the same ear. It can be shown that

FIG. 26. Masking audiogram for street noise.

nerve deafness is equivalent to having a noise continually present which produces a similar masking of sounds, while obstructive deafness is equivalent to introducing attenuation between the source of the sound and the listener. If we unroll one of these auditory patterns and make it into an ordinary rectangular plot, for example one corresponding to pure tones of 200 cycles and 2000 cycles, we will get the curves shown at the top of Fig. 29.

Now if we change the two scales, plotting masking in decibels as ordinates instead of loudness per patch and plotting frequency as abscissae instead of nerve position x, then the top figure resolves itself into the one shown at the bottom of Fig. 29. This plot as it now stands is called the noise audiogram and it gives the experimental data as they are usually tabulated. It has been shown that it is such noise audiograms which give the fundamental data from which the auditory patterns can be drawn. This plot differs from those shown earlier in that the frequency is plotted on a logarithmic rather than a linear frequency scale. This makes it possible to plot the audible range of frequencies in a single diagram.

Similarly, when a person is deafened we represent the amount of the deafness by this same type of audiogram, but of course in this case there is no noise present and the ordinates represent the decibels that tones of the frequencies indicated by the abscissae must be raised above that for normal hearing. So we can say, that a person of normal hearing who is immersed in a noise which produces a given

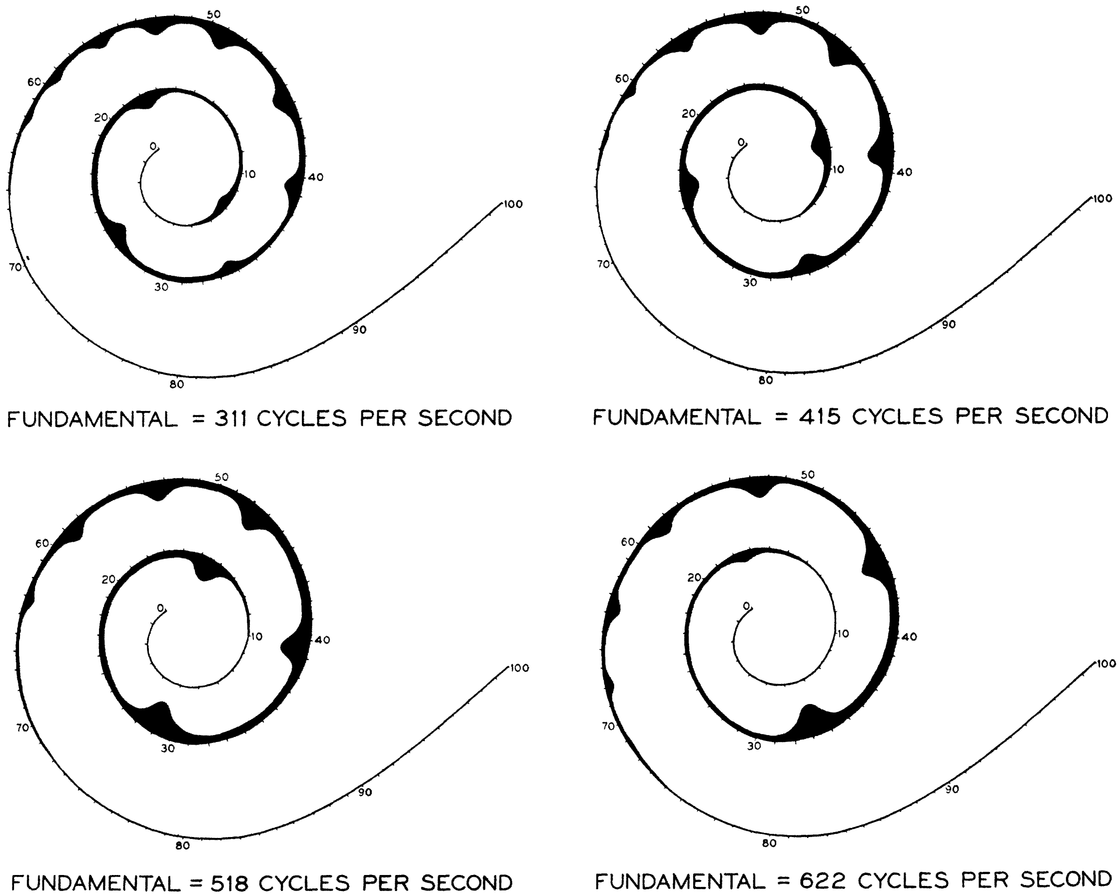

FIG. 28. Auditory patterns for four notes of bugle playing "taps."

audiogram will have only those capabilities for hearing desired sounds that a deafened person would have with the same kind of a deafness audiogram. Because of this equivalence a person of normal hearing who is immersed in a noise experiences the same difficulties of hearing speech which are actually experienced by deafened persons.

Let us consider a moderate case of nerve deafness. The audiogram for such a case is shown in Fig. 30, curve A. If we use this audiogram and plot the corresponding auditory pattern, we get the results shown in Fig. 31. The breadth of the crosshatched portion now indicates the amount of damage to the nerve endings at the various positions along the basilar membrane. Instead of an auditory pattern I will call this area a deafness pattern. The scale is so chosen that if this deafened person is to hear any sound the pattern which would be plotted for a normal hearing person must project through the pattern

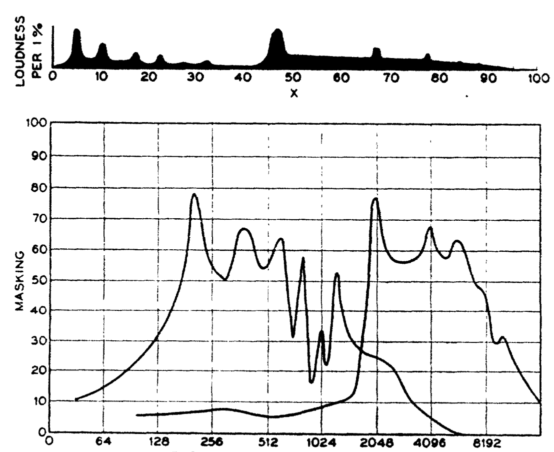

FIG. 29. Upper curve, auditory pattern on a rectangular plot. Lower curve, masking in decibels *vs.* frequency.

here shown. In other words, the normal auditory pattern must project through the deafness pattern. For example, in Fig. 30, curve B, is shown the audiogram for a case of nerve deafness which is 30 db more severe than the one in Fig. 30, curve A. The corresponding deafness pattern is shown by the crosshatched portions of Fig. 32.

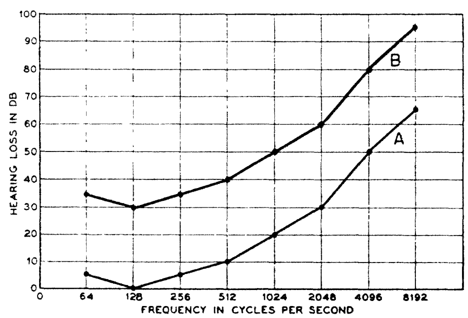

FIG. 30. Nerve deafness audiograms.

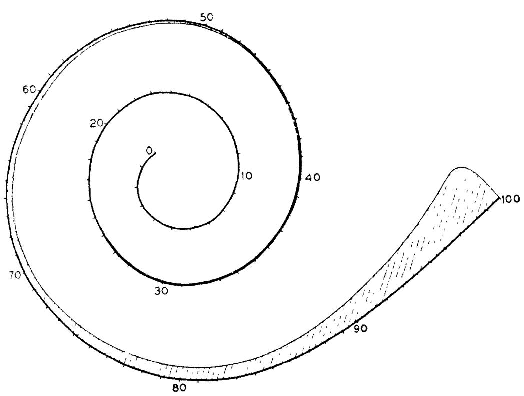

FIG. 31. Nerve deafness pattern.

Also, on this deafness pattern are shown, by the white portions, the normal auditory patterns produced by the pure tones having frequencies of 2300 cycles and 6500 cycles. It will be seen that the latter tone will not be heard because its normal auditory pattern does not project through the crosshatched area corresponding to the deafness pattern. However, a 2300-cycle tone will be heard and the loudness with which it will be heard is equal to the amount of whitened area projecting above the crosshatched area in this figure. Indeed, loudness tests upon such deafened persons have verified that these results agree with such calculated values of loudness. So the auditory pattern for a person having

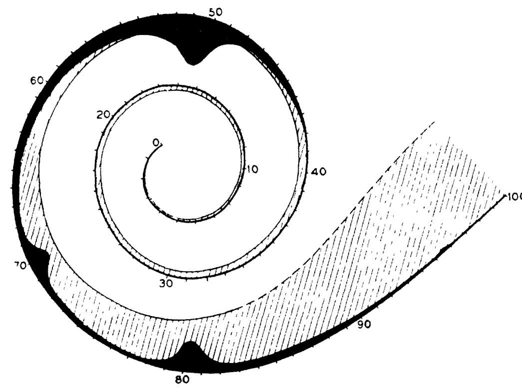

FIG. 32. Nerve deafness pattern.

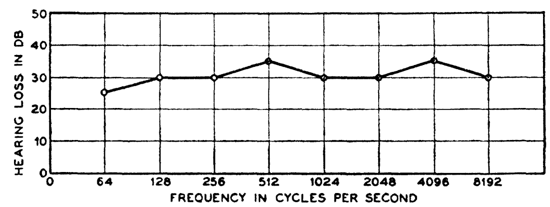

FIG. 33. Obstructive deafness audiogram.

nerve deafness is that part of the normal auditory pattern which protrudes above the deafness pattern. It has been shown that both the normal auditory pattern and also the deafness pattern can be obtained quantitatively from audition measurements. Consequently the auditory pattern for the deafened person can also be drawn in a quantitative way.[12]

Obstructive deafness is equivalent to putting attenuation in the transmission path between the source of sound and the listener. In Fig. 33 is shown the audiogram for a typical case of obstructive deafness. Now we wish to know what the auditory pattern is for such a deafened person when he listens to the steamboat whistle whose audiogram was shown in Fig. 25. In Fig. 34 the deafness audiogram is shown again by the line at the bottom and the noise audiogram for the steamboat whistle is shown by the top heavy line. Since obstructive deafness is equivalent to attenuation between the source of sound and the listener, then the effective noise audiogram for this deafened person is just the difference between the ordinates of these curves, or that

[12] See paper by Steinberg and Gardner, J. Acous. Soc. Am. 9, 11 (1937).

given by the dotted line. This effective noise audiogram can now be used to plot the corresponding auditory pattern in the manner outlined above and it will correspond to the pattern of sensation experienced by such a deafened person. Such an auditory pattern for the steamboat whistle listened to by a person deafened by the amount shown in Fig. 33 is given in Fig. 35. The loudness experienced by such a deafened person is equal to the area of this auditory pattern. Loudness tests with persons having such obstructive deafness agree with these calculated results.

To treat the case of mixed deafness we must separate the deafness audiogram into two parts, one due to obstructive deafness and the other due to nerve deafness. This can be done by finding by tests a deafness audiogram by air conduction (the audiogram used above) and then by bone conduction. In the bone conduction test the testing instrument conducts the sound directly into the bones of the head. The vibration thus produced by-passes the transmission mechanism of the middle ear and is transmitted directly to the fluid of the inner ear. Consequently, a bone conduction audiogram gives the hearing loss due to nerve deafness, or the nerve

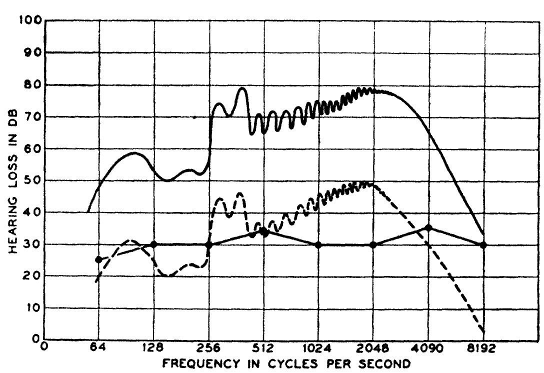

FIG. 34. Noise audiogram of steamboat whistle (top line) and the deafness audiogram (bottom solid line) for a case of obstructive deafness.

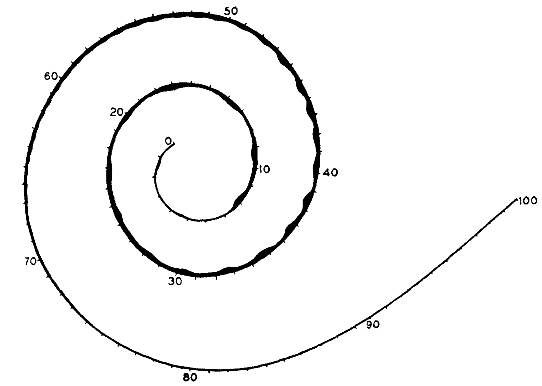

FIG. 35. Auditory pattern for steamboat whistle—for deafened listener.

deafness audiogram. The difference of the ordinates in the bone conduction audiogram and the air conduction audiogram gives the amount of obstructive deafness, or the obstructive deafness audiogram. Having these two audiograms, the procedure is as follows. First apply the obstructive deafness audiogram to the noise audiogram, subtracting the ordinates of the former from the latter. Then use the difference audiogram to construct the auditory pattern of the noise. Next construct the nerve deafness pattern from the nerve deafness audiogram. The part of the difference auditory pattern above the nerve deafness auditory pattern is the pattern which we are seeking and corresponds to that experienced by the person having the mixed deafness.

It has been shown that an auditory pattern shows the amount and kind of sensation being transmitted from the inner ear to the brain. To determine it for any person A immersed in a sound B, one needs to measure, first, the noise audiogram of the sound B for a typical normal ear, and second, the obstructive and nerve deafness audiograms for the person A. Then from the relations established in this paper, the auditory pattern can be drawn for the person A listening to the sound B.

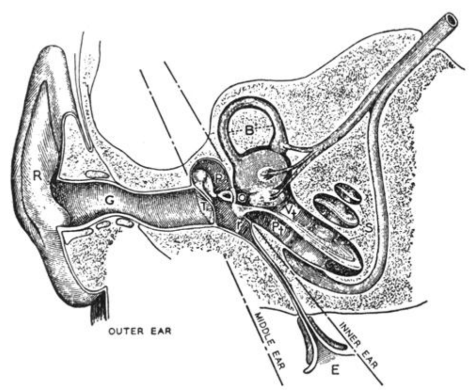

FIG. 5. Semi-diagrammatic section through the right ear (Czermak). *G*—external auditory meatus; *T*—membrana tympani; *P*—tympanic cavity; *o*—fenestra ovalis; *r*—fenestra rotunda; *B*—semi-circular canal; *S*—cochlea; *Vt*—scala vestibuli; *Pt*—scala tympani; *E*—Eustachian tube; *R*—pinna.