The Bell System Technical Journal

October, 1933

Loudness, Its Definition, Measurement and Calculation*

By HARVEY FLETCHER and W. A. MUNSON

An empirical formula for calculating the loudness of any steady sound from an analysis of the intensity and frequency of its components is developed in this article. The development is based on fundamental properties of the hearing mechanism in such a way that a scale of loudness values results. In order to determine the form of the function representing this loudness scale and of the other factors entering into the loudness formula, measurements were made of the loudness levels of many sounds, both of pure tones and of complex wave forms. These tests are described and the method of measuring loudness levels is discussed in detail. Definitions are given endeavoring to clarify the terms used and the measurement of the physical quantities which determine the characteristics of a sound wave stimulating the auditory mechanism.

Introduction

LOUDNESS is a psychological term used to describe the magnitude of an auditory sensation. Although we use the terms "very loud," "loud," "moderately loud," "soft" and "very soft," corresponding to the musical notations *ff*, *f*, *mf*, *p*, and *pp*, to define the magnitude, it is evident that these terms are not at all precise and depend upon the experience, the auditory acuity, and the customs of the persons using them. If loudness depended only upon the intensity of the sound wave producing the loudness, then measurements of the physical intensity would definitely determine the loudness as sensed by a typical individual and therefore could be used as a precise means of defining it. However, no such simple relation exists.

The magnitude of an auditory sensation, that is, the loudness of the sound, is probably dependent upon the total number of nerve impulses that reach the brain per second along the auditory tract. It is evident that these auditory phenomena are dependent not alone upon the intensity of the sound but also upon their physical composition. For example, if a person listened to a flute and then to a bass drum placed at such distances that the sounds coming from the two instruments are judged to be equally loud, then the intensity of the sound at the ear produced by the bass drum would be many times that produced by the flute.

If the composition of the sound, that is, its wave form, is held constant, but its intensity at the ear of the listener varied, then the loud-

* Jour. Acous. Soc. Amer., October, 1933.

ness produced will be the same for the same intensity only if the same or an equivalent ear is receiving the sound and also only if the listener is in the same psychological and physiological conditions, with reference to fatigue, attention, alertness, etc. Therefore, in order to determine the loudness produced, it is necessary to define the intensity of the sound, its physical composition, the kind of ear receiving it, and the physiological and psychological conditions of the listener. In most engineering problems we are interested mainly in the effect upon a typical observer who is in a typical condition for listening.

In a paper during 1921 one of us suggested using the number of decibels above threshold as a measure of loudness and some experimental data were presented on this basis. As more data were accumulated it was evident that such a basis for defining loudness must be abandoned.

In 1924 in a paper by Steinberg and Fletcher [1] some data were given which showed the effects of eliminating certain frequency bands upon the loudness of the sound. By using such data as a basis, a mathematical formula was given for calculating the loudness losses of a sound being transmitted to the ear, due to changes in the transmission system. The formula was limited in its application to the particular sounds studied, namely, speech and a sound which was generated by an electrical buzzer and called the test tone.

In 1925 Steinberg [2] developed a formula for calculating the loudness of any complex sound. The results computed by this formula agreed with the data which were then available. However, as more data have accumulated it has been found to be inadequate. Since that time considerably more information concerning the mechanism of hearing has been discovered and the technique in making loudness measurements has advanced. Also more powerful methods for producing complex tones of any known composition are now available. For these reasons and because of the demand for a loudness formula of general application, especially in connection with noise measurements, the whole subject was reviewed by the Bell Telephone Laboratories and the work reported in the present paper undertaken. This work has resulted in better experimental methods for determining the loudness level of any sustained complex sound and a formula which gives calculated results in agreement with the great variety of loudness data which are now available.

[1] H. Fletcher and J. C. Steinberg, "Loudness of a Complex Sound," *Phys, Rev.* **24,** 306 (1924).
[2] J. C. Steinberg, "The Loudness of a Sound and Its Physical Stimulus," *Phys. Rev.* **26,** 507 (1925).

DEFINITIONS

The subject matter which follows necessitates the use of a number of terms which have often been applied in very inexact ways in the past. Because of the increase in interest and activity in this field, it became desirable to obtain a general agreement concerning the meaning of the terms which are most frequently used. The following definitions are taken from recent proposals of the sectional committee on Acoustical Measurements and Terminology of the American Standards Association and the terms have been used with these meanings throughout the paper.

Sound Intensity

The sound intensity of a sound field in a specified direction at a point is the sound energy transmitted per unit of time in the specified direction through a unit area normal to this direction at the point.

In the case of a plane or spherical free progressive wave having the effective sound pressure P (bars), the velocity of propagation c (cm. per sec.) in a medium of density ρ (grams per cubic cm.), the intensity in the direction of propagation is given by

$$J = P^2/\rho c \text{ (ergs per sec. per sq. cm.)}. \tag{1}$$

This same relation can often be used in practice with sufficient accuracy to calculate the intensity at a point near the source with only a pressure measurement. In more complicated sound fields the results given by this relation may differ greatly from the actual intensity.

When dealing with a plane or a spherical progressive wave it will be understood that the intensity is taken in the direction of propagation of the wave.

Reference Intensity

The reference intensity for intensity level comparisons shall be 10^{-16} watts per square centimeter. In a plane or spherical progressive sound wave in air, this intensity corresponds to a root-mean-square pressure p given by the formula

$$p = 0.000207[(H/76)(273/T)^{\frac{1}{2}}]^{\frac{1}{2}} \tag{2}$$

where p is expressed in bars, H is the height of the barometer in centimeters, and T is the absolute temperature. At a temperature of 20° C. and a pressure of 76 cm. of Hg, $p = 0.000204$ bar.

Intensity Level

The intensity level of a sound is the number of db above the reference intensity.

Reference Tone

A plane or spherical sound wave having only a single frequency of 1,000 cycles per second shall be used as the reference for loudness comparisons.

Note: One practical way to obtain a plane or spherical wave is to use a small source, and to have the head of the observer at least one meter distant from the source, with the external conditions such that reflected waves are negligible as compared with the original wave at the head of the observer.

Loudness Level

The loudness level of any sound shall be the intensity level of the equally loud reference tone at the position where the listener's head is to be placed.

Manner of Listening to the Sound

In observing the loudness of the reference sound, the observer shall face the source, which should be small, and listen with both ears at a position so that the distance from the source to a line joining the two ears is one meter.

The value of the intensity level of the equally loud reference sound depends upon the manner of listening to the unknown sound and also to the standard of reference. The manner of listening to the unknown sound may be considered as part of the characteristics of that sound. The manner of listening to the reference sound is as specified above.

Loudness has been briefly defined as the magnitude of an auditory sensation, and more will be said about this later, but it will be seen from the above definitions that the *loudness level* of any sound is obtained by adjusting the intensity level of the reference tone until it sounds equally loud as judged by a typical listener. The only way of determining a typical listener is to use a number of observers who have normal hearing to make the judgment tests. The typical listener, as used in this sense, would then give the same results as the average obtained by a large number of such observers.

A pure tone having a frequency of 1000 cycles per second was chosen for the reference tone for the following reasons: (1) it is simple to define, (2) it is sometimes used as a standard of reference for pitch, (3) its use makes the mathematical formulae more simple, (4) its range of auditory intensities (from the threshold of hearing to the threshold of feeling) is as large and usually larger than for any other type of sound, and (5) its frequency is in the mid-range of audible frequencies.

There has been considerable discussion concerning the choice of the reference or zero for loudness levels. In many ways the threshold of hearing intensity for a 1000-cycle tone seems a logical choice. However, variations in this threshold intensity arise depending upon the individual, his age, the manner of listening, the method of presenting the tone to the listener, etc. For this reason no attempt was made to choose the reference intensity as equal to the average threshold of a given group listening in a prescribed way. Rather, an intensity of the reference tone in air of 10^{-16} watts per square centimeter was chosen as the reference intensity because it was a simple number which was convenient as a reference for computation work, and at the same time it is in the range of threshold measurements obtained when listening in the standard method described above. This reference intensity corresponds to the threshold intensity of an observer who might be designated a reference observer. An examination of a large series of measurements on the threshold of hearing indicates that such a reference observer has a hearing which is slightly more acute than the average of a large group. For those who have been thinking in terms of microwatts it is easy to remember that this reference level is 100 db below one microwatt per square centimeter. When using these definitions the intensity level β_r of the reference tone is the same as its loudness level L and is given by

$$\beta_r = L = 10 \log J_r + 100, \qquad (3)$$

where J_r is its sound intensity in microwatts per square centimeter.

The intensity level of any other sound is given by

$$\beta = 10 \log J + 100, \qquad (4)$$

where J is its sound intensity, but the loudness level of such a sound is a complicated function of the intensities and frequencies of its components. However, it will be seen from the experimental data given later that for a considerable range of frequencies and intensities the intensity level and loudness level for pure tones are approximately equal.

With the reference levels adopted here, all values of loudness level which are positive indicate a sound which can be heard by the reference observer and those which are negative indicate a sound which cannot be heard by such an observer.

It is frequently more convenient to use two matched head receivers for introducing the reference tone into the two ears. This can be done provided they are calibrated against the condition described above. This consists in finding by a series of listening tests by a number of observers the electrical power W_1 in the receivers which produces the same loudness as a level β_1 of the reference tone. The intensity level β_r of an open air reference tone equivalent to that produced in the receiver for any other power W_r in the receivers is then given by

$$\beta_r = \beta_1 + 10 \log (W_r/W_1). \tag{5}$$

Or, since the intensity level β_r of the reference tone is its loudness level L, we have

$$L = 10 \log W_r + C_r, \tag{6}$$

where C_r is a constant of the receivers.

In determining loudness levels by comparison with a reference tone there are two general classes of sound for which measurements are desired: (1) those which are steady, such as a musical tone, or the hum from machinery, (2) those which are varying in loudness such as the noise from the street, conversational speech, music, etc. In this paper we have confined our discussion to sources which are steady and the method of specifying such sources will now be given.

A steady sound can be represented by a finite number of pure tones called components. Since changes in phase produce only second order effects upon the loudness level it is only necessary to specify the magnitude and frequency of the components.[3] The magnitudes of the components at the listening position where the loudness level is desired are given by the intensity levels $\beta_1, \beta_2, \cdots \beta_k, \cdots \beta_n$ of each component at that position. In case the sound is conducted to the ears by telephone receivers or tubes, then a value W_k for each component must be known such that if this component were acting separately it would produce the same loudness for typical observers as a tone of the same pitch coming from a source at one meter's distance and producing an intensity level of β_k.

In addition to the frequency and magnitude of the components of a sound it is necessary to know the position and orientation of the head with respect to the source, and also whether one or two ears are used in listening. The monaural type of listening is important in telephone use and the binaural type when listening directly to a sound source in air. Unless otherwise stated, the discussion and data which follow apply to the condition where the listener faces the source and uses both ears, or uses head telephone receivers which produce an equivalent result.

[3] Recent work by Chapin and Firestone indicates that at very high levels these second order effects become large and cannot be neglected. K. E. Chapin and F. A. Firestone, "Interference of Subjective Harmonics," *Jour. Acous. Soc. Am.* **4**, 176*A* (1933).

Formulation of the Empirical Theory for Calculating the Loudness Level of a Steady Complex Tone

It is well known that the intensity of a complex tone is the sum of the intensities of the individual components. Similarly, in finding a method of calculating the loudness level of a complex tone one would naturally try to find numbers which could be related to each component in such a way that the sum of such numbers will be related in the same way to the equally loud reference tone. Such efforts have failed because the amount contributed by any component toward the total loudness sensation depends not only upon the properties of this component but also upon the properties of the other components in the combination. The answer to the problem of finding a method of calculating the loudness level lies in determining the nature of the ear and brain as measuring instruments in evaluating the magnitude of an auditory sensation.

One can readily estimate roughly the magnitude of an auditory sensation; for example, one can tell whether the sound is soft or loud. There have been many theories to account for this change in loudness. One that seems very reasonable to us is that the loudness experienced is dependent upon the total number of nerve impulses per second going to the brain along all the fibers that are excited. Although such an assumption is not necessary for deriving the formula for calculating loudness it aids in making the meaning of the quantities involved more definite.

Let us consider, then, a complex tone having n components each of which is specified by a value of intensity level β_k and of frequency f_k. Let N be a number which measures the magnitude of the auditory sensation produced when a typical individual listens to a pure tone. *Since by definition the magnitude of an auditory sensation is the loudness, then N is the loudness of this simple tone.* Loudness as used here must not be confused with loudness level. The latter is measured by the intensity of the equally loud reference tone and is expressed in decibels while the former will be expressed in units related to loudness levels in a manner to be developed. If we accept the assumption mentioned above, N is proportional to the number of nerve impulses per second reaching the brain along all the excited nerve fibers when the typical observer listens to a simple tone.

Let the dependency of the loudness N upon the frequency f and the intensity β for a simple tone be represented by

$$N = G(f, \beta), \tag{7}$$

where G is a function which is determined by any pair of values of f

and β. For the reference tone, f is 1000 and β is equal to the loudness level L, so a determination of the relation expressed in Eq. (7) for the reference tone gives the desired relation between loudness and loudness level.

If now a simple tone is put into combination with other simple tones to form a complex tone, its loudness contribution, that is, its contribution toward the total sensation, will in general be somewhat less because of the interference of the other components. For example, if the other components are much louder and in the same frequency region the loudness of the simple tone in such a combination will be zero. Let $1 - b$ be the fractional reduction in loudness because of its being in such a combination. Then bN is the contribution of this component toward the loudness of the complex tone. It will be seen that b by definition always remains between 0 and unity. It depends not only upon the frequency and intensity of the simple tone under discussion but also upon the frequencies and intensities of the other components. It will be shown later that this dependence can be determined from experimental measurements.

The subscript k will be used when f and β correspond to the frequency and intensity level of the kth component of the complex tone, and the subscript r used when f is 1000 cycles per second. The "loudness level" L by definition, is the intensity level of the reference tone when it is adjusted so it and the complex tone sound equally loud. Then

$$N_r = G(1000, L) = \sum_{k=1}^{k=n} b_k N_k = \sum_{k=1}^{k=n} b_k G(f_k, \beta_k). \tag{8}$$

Now let the reference tone be adjusted so that it sounds equally loud successively to simple tones corresponding in frequency and intensity to each component of the complex tone.

Designate the experimental values thus determined as $L_1, L_2, L_3, \cdots L_k, \cdots L_n$. Then from the definition of these values

$$N_k = G(1000, L_k) = G(f_k, \beta_k), \tag{9}$$

since for a single tone b_k is unity. On substituting the values from (9) into (8) there results the fundamental equation for calculating the loudness of a complex tone

$$G(1000, L) = \sum_{k=1}^{k=n} b_k G(1000, L_k). \tag{10}$$

This transformation looks simple but it is a very important one since instead of having to determine a different function for every component, we now have to determine a single function depending only upon the properties of the reference tone and as stated above this function is the relationship between loudness and loudness level. And since the frequency is always 1000 this function is dependent only upon the single variable, the intensity level.

This formula has no practical value unless we can determine b_k and G in terms of quantities which can be obtained by physical measurements. It will be shown that experimental measurements of the loudness levels L and L_k upon simple and complex tones of a properly chosen structure have yielded results which have enabled us to find the dependence of b and G upon the frequencies and intensities of the components. When b and G are known, then the more general function $G(f, \beta)$ can be obtained from Eq. (9), and the experimental values of L_k corresponding to f_k and β_k.

Determination of the Relation Between L_k, f_k and β_k

This relation can be obtained from experimental measurements of the loudness levels of pure tones. Such measurements were made by Kingsbury [4] which covered a range in frequency and intensity limited by instrumentalities then available. Using the experimental technique described in Appendix A, we have again obtained the loudness levels of pure tones, this time covering practically the whole audible range. (See Appendix B for a comparison with Kingsbury's results.)

All of the data on loudness levels both for pure and also complex tones taken in our laboratory which are discussed in this paper have been taken with telephone receivers on the ears. It has been explained previously how telephone receivers may be used to introduce the reference tone into the ears at known loudness levels to obtain the loudness levels of other sounds by a loudness balance. If the receivers are also used for producing the sounds whose loudness levels are being determined, then an additional calibration, which will be explained later, is necessary if it is desired to know the intensity levels of the sounds.

The experimental data for determining the relation between L_k and f_k are given in Table I in terms of voltage levels. (Voltage level $= 20 \log V$, where V is the e.m.f. across the receivers in volts.) The pairs of values in each double column give the voltage levels of the reference tone and the pure tone having the frequency indicated at the top of the column when the two tones coming from the head receivers were judged to be equally loud when using the technique

[4] B. A. Kingsbury, "A Direct Comparison of the Loudness of Pure Tones," *Phys. Rev.* **29**, 588 (1927).

TABLE I
Voltage Levels (db) for Loudness Equality

Reference	62 c.p.s.
−12.2	+ 9.8
−17.2	+ 5.8
−19.2	+ 2.8
−15.7	+ 2.6
−21.2	+ 0.8
−27.2	− 0.2
−32.2	− 7.2
−33.2	− 7.2
−41.2	−10.2
−35.4	−10.4
−56.2	−15.2
−67.2	−20.2
−68.7	−20.3
−97.2	−30.3
−108.1	−39.8
−108.3	−39.5
−109.3	−42.4
−113.1	−38.5

Reference	125 c.p.s.
− 4.4	+ 9.8
−10.2	+ 7.9
−13.3	− 0.8
−18.6	− 3.2
−23.2	− 5.2
−27.9	−12.3
−31.0	−14.2
−35.2	−15.2
−40.7	−23.6
−66.6	−35.0
−88.8	−46.5
−108.1	−62.8
−108.3	−60.7
−113.1	−63.5

Reference	250 c.p.s.
− 2.9	+ 6.6
− 3.7	+ 5.7
− 5.2	+ 5.8
− 6.7	− 2.2
−12.2	− 2.2
−25.5	−18.3
−32.2	−22.2
−32.2	−23.2
−52.5	−40.4
−72.9	−56.3
−90.2	−68.3
−108.1	−86.7
−108.3	−86.4
−113.1	−86.3

Reference	500 c.p.s.
− 2.2	+ 5.8
− 4.2	+ 6.8
− 6.2	− 2.0
− 7.2	− 2.3
−12.2	− 8.2
−21.2	−22.3
−21.7	−21.9
−32.2	−30.2
−34.2	−31.2
−41.7	−41.9
−43.7	−42.2
−63.7	−61.0
−64.2	−61.2
−83.2	−80.2
−83.7	−78.0
−108.1	−102.6
−108.3	−101.7
−108.3	− 99.7
−109.3	−103.4
−113.1	−103.0

Reference	2000 c.p.s.
− 2.2	− 1.2
− 1.7	− 1.5
− 3.2	− 1.3
−18.2	−13.4
−21.2	−17.2
−22.2	−18.3
−40.2	−35.2
−41.2	−35.3
−42.2	−35.4
−59.2	−54.2
−61.2	−57.5
−64.2	−57.3
−78.2	−77.3
−78.2	−80.2
−81.7	−77.3
−108.3	−105.2
−108.3	−109.0
−109.3	−108.9
−113.1	−111.4

Reference	4000 c.p.s.
− 1.7	− 0.7
− 1.2	− 1.5
−23.2	−17.3
−24.7	−16.7
−44.7	−35.2
−47.7	−35.1
−63.2	−54.2
−65.2	−54.5
−77.7	−72.2
−80.2	−72.5
−108.3	−104.6
−108.3	−105.7
−109.3	−102.0
−113.1	−108.1

Reference	5650 c.p.s.
− 3.7	+ 0.8
− 7.2	− 1.2
−28.2	−19.2
−30.2	−19.2
−53.2	−38.2
−54.7	−39.1
−72.2	−58.2
−72.7	−58.1
−85.2	−71.2
−92.7	−78.1
−108.3	−101.9
−109.3	− 99.3
−113.1	−102.3

Reference	8000 c.p.s.
−10.9	− 4.3
−12.2	−12.0
−26.2	−24.0
−27.1	−24.3
−46.2	−38.3
−48.2	−38.2
−64.2	−52.2
−70.2	−56.2
−76.2	−72.2
−82.6	−76.2
−108.3	−108.1
−109.3	−103.1
−113.1	−106.6

Reference	11,300 c.p.s.
−20.2	+ 1.7
−22.2	+ 1.8
−38.2	−20.2
−38.7	−20.3
−55.2	−34.2
−55.7	−34.2
−72.2	−52.2
−78.7	−58.1
−88.2	−72.2
−90.7	−77.1
−108.3	−93.7
−109.3	−94.6
−113.1	−93.7

Reference	16,000 c.p.s.
−50.2	+ 1.8
−66.2	−13.2
−77.2	−28.2
−85.2	−38.2
−109.3	−57.3

described in Appendix A. For example, in the second column it will be seen that for the 125-cycle tone when the voltage is + 9.8 db above 1 volt then the voltage level for the reference tone must be 4.4 db below 1 volt for equality of loudness. The bottom set of numbers in each column gives the threshold values for this group of observers.

Each voltage level in Table I is the median of 297 observations representing the combined results of eleven observers. The method of obtaining these is explained in Appendix A also. The standard deviation was computed and it was found to be somewhat larger for tests in which the tone differed most in frequency from the reference tone. The probable error of the combined result as computed in the usual way was between 1 and 2 db. Since deviations of any one observer's results from his own average are less than the deviations of his average from the average of the group, it would be necessary to increase the size of the group if values more representative of the average normal ear were desired.

The data shown in Table I can be reduced to the number of decibels above threshold if we accept the values of this crew as the reference threshold values. However, we have already adopted a value for the 1000-cycle reference zero. As will be shown, our crew obtained a threshold for the reference tone which is 3 db above the reference level chosen.

It is not only more convenient but also more reliable to relate the data to a calibration of the receivers in terms of physical measurements of the sound intensity rather than to the threshold values. Except in experimental work where the intensity of the sound can be definitely controlled, it is obviously impractical to measure directly the threshold level by using a large group of observers having normal hearing. For most purposes it is more convenient to measure the intensity levels $\beta_1, \beta_2, \cdots \beta_k$, etc., directly rather than have them related in any way to the threshold of hearing.

In order to reduce the data in Table I to those which one would obtain if the observers were listening to a free wave and facing the source, we must obtain a field calibration of the telephone receivers used in the loudness comparisons. The calibration for the reference tone frequency has been explained previously and the equation

$$\beta_r = \beta_1 + 10 \log (W_r/W_1) \tag{5}$$

derived for the relation between the intensity β_r of the reference tone and the electrical power W_r in the receivers. The calibration consisted of finding by means of loudness balances a power W_1 in the receivers which produces a tone equal in loudness to that of a free wave having an intensity level β_1.

For sounds other than the 1000-cycle reference tone a relation similar to Eq. (5) can be derived, namely,

$$\beta = \beta_1 + 10 \log (W/W_1), \tag{11}$$

where β_1 and W_1 are corresponding values found from loudness balances for each frequency or complex wave form of interest. If, as is usually assumed, a linear relation exists between β and 10 log W, then determinations of β_1 and W_1 at one level are sufficient and it follows that a change in the power level of Δ decibels will produce a corresponding change of Δ decibels in the intensity of the sound generated. Obviously the receivers must not be overloaded or this assumption will not be valid. Rather than depend upon the existence of a linear relation between β and 10 log W with no confirming data, the receivers used in this investigation were calibrated at two widely separated levels.

Referring again to Table I, the data are expressed in terms of voltage levels instead of power levels. If, as was the case with our receivers, the electrical impedance is essentially a constant, Eq. (11) can be put in the form:

$$\beta = \beta_1 + 20 \log (V/V_1) \tag{12}$$

or

$$\beta = 20 \log V + C, \tag{13}$$

where V is the voltage across the receivers and C is a constant of the receivers to be determined from a calibration giving corresponding values of β_1 and 20 log V_1. The calibration will now be described.

By using the sound stage and the technique of measuring field pressures described by Sivian and White [5] and by using the technique for making loudness measurements described in Appendix A, the following measurements were made. An electrical voltage V_1 was placed across the two head receivers such that the loudness level produced was the same at each frequency. The observer listened to the tone in these head receivers and then after $1\frac{1}{2}$ seconds silence listened to the tone from the loud speaker producing a free wave of the same frequency. The voltage level across the loud speaker necessary to produce a tone equally loud to the tone from the head receivers was obtained using the procedure described in Appendix A. The free wave intensity level β_1 corresponding to this voltage level was measured in the manner described in Sivian and White's paper. Threshold values both for the head receivers and the loud speaker were also observed. In these tests eleven observers were used. The results obtained are given in Table II. In the second row values of 20 log V_1, the voltage

[5] L. J. Sivian and S. D. White, "Minimum Audible Sound Fields," *Jour. Acous. Soc. Am.* **4**, 288 (1933).

TABLE II
FIELD CALIBRATION OF TELEPHONE RECEIVERS

Frequency c.p.s.	60	120	240	480	960	1920	3850	5400	7800	10,500	15,000
Voltage level (20 log V_1)	−13.0	−26.2	−38.5	−47.0	−48.2	−42.3	−36.3	−34.0	−39.1	−32.4	− 6.4
Intensity level (β_1)	+79.3	+71.0	+67.4	+63.8	+65.3	+64.0	+62.2	+65.5	+74.0	+78.6	+75.0
$C_1 = \beta_1 = -20 \log V_1$	92.3	97.2	105.9	110.8	113.5	106.3	98.5	99.5	113.1	111.0	81.4
Threshold voltage level (20 log V_0)	−48.0	−61.8	−86.2	−105.4	−110.7	−109.0	−104.0	−97.1	−100.5	−102.0	−74.0
Threshold intensity level (β_0)	+49.3	+33.7	+19.7	+8.4	+5.4	−0.9	−4.2	+2.7	+10.6	+16.1	+22.0
$C_0 = \beta_0 - 20 \log V_0$	97.3	95.5	105.9	113.8	116.1	108.1	99.8	99.8	111.1	118.1	96.0
Diff. $= C_1 - C_0$	−5.0	1.7	0	−3.0	−2.6	−1.8	−1.3	−0.3	+2.0	−7.1	−14.6

level, are given. The intensity levels, β_1, of the free wave which sounded equally loud are given in the third row. In the fourth row the values of the constant C, the calibration we are seeking, are given. The voltage level added to this constant gives the equivalent free wave intensity level. In the fifth, sixth and seventh rows, similar values are given which were determined at the threshold level. In the bottom row the differences in the constants determined at the two levels are given. The fact that the difference is no larger than the probable error is very significant. It means that throughout this wide range there is a linear relationship between the equivalent field intensity levels, β, and the voltage levels, 20 log V, so that the formula (13)

$$\beta = 20 \log V + C$$

can be applied to our receivers with considerable confidence.

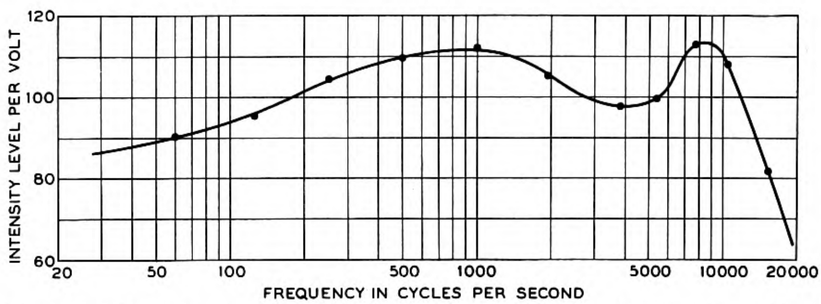


Fig. 1—Field calibration of loudness balance receivers.[6] (Calibration made at $L = 60$ db.)

The constant C determined at the high level was determined with greater accuracy than at the threshold. For this reason only the values for the higher level were used for the calibration curve. Also in these tests only four receivers were used while in the loudness tests eight receivers were used. The difference between the efficiency of the former four and the latter eight receivers was determined by measurements on an artificial ear. The figures given in Table II were corrected by this difference. The resulting calibration curve is that given in Fig. 1. It should be pointed out here that such a calibration curve on a single individual would show considerable deviations from this average curve. These deviations are real, that is, they are due to the sizes and shapes of the ear canals.

[6] The ordinates represent the intensity level in db of a free wave in air which, when listened to with both ears in the standard manner, is as loud as a tone of the same frequency heard from the two head receivers used in the tests when an e.m.f. of one volt is applied to the receiver terminals.

We can now express the data in Table I in terms of field intensity levels. To do this, the data in each double column were plotted and a smooth curve drawn through the observed points. The resulting curves give the relation between voltage levels of the pure tones for equality of loudness. From the calibration curve of the receivers these levels are converted to intensity levels by a simple shift in the axes of coordinates. Since the intensity level of the reference tone is by definition the "loudness level," these shifted curves will represent the loudness level of pure tones in terms of intensity levels. The resulting curves for the ten tones tested are given in Figs. 2A to 2J. Each point on these curves corresponds to a pair of values in Table I except for the threshold values. The results of separate determinations by the crew used in these loudness tests at different times are given by the circles. The points represented by (*) are the values adopted by Sivian and White. It will be seen that most of the

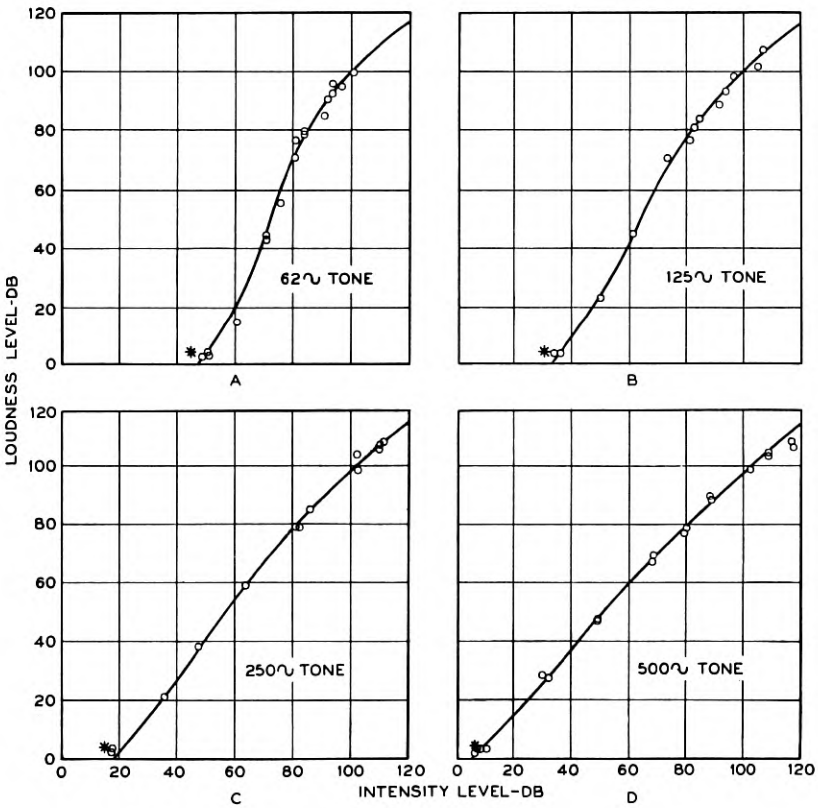


Fig. 2 (A to D)—Loudness levels of pure tones.

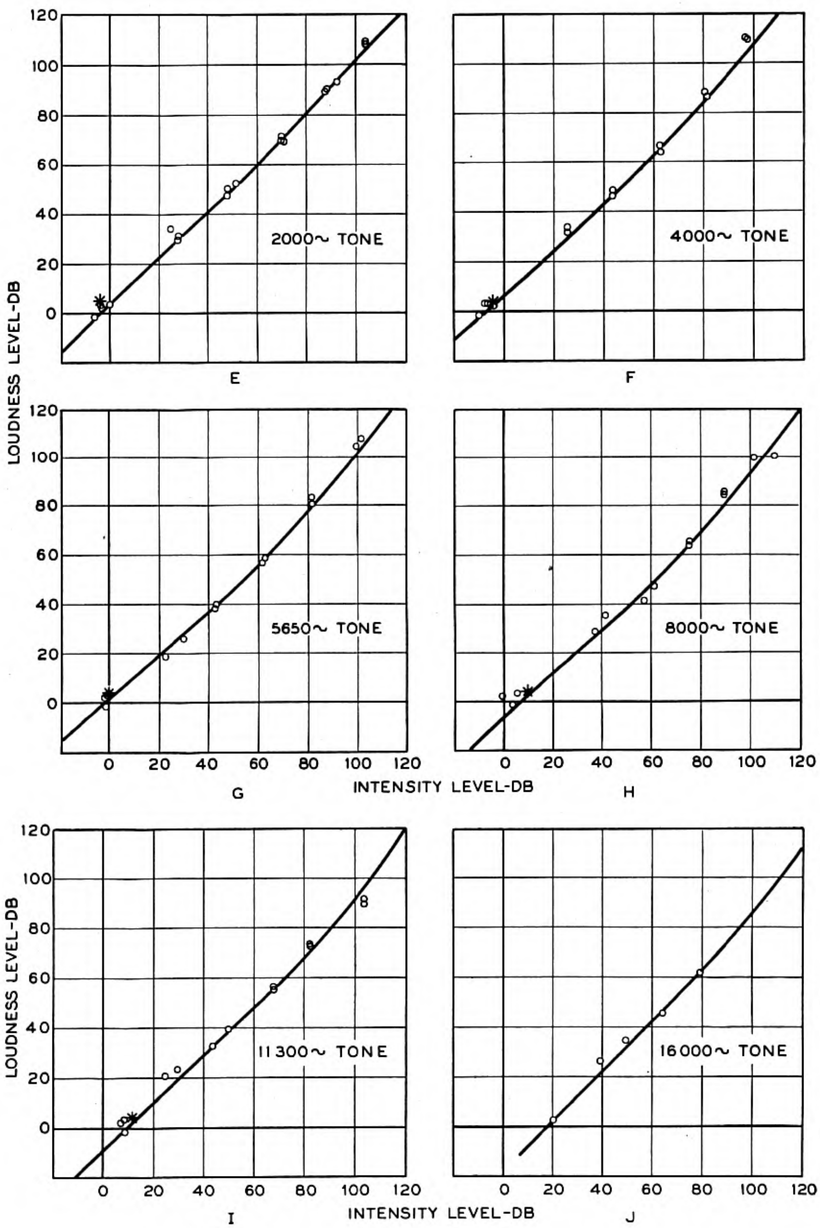


Fig. 2 (E to J)—Loudness levels of pure tones.

threshold points are slightly above the zero we have chosen. This means that our zero corresponds to the thresholds of observers who are slightly more acute than the average.

From these curves the loudness level contours can be drawn. The first set of loudness level contours are plotted with levels above reference threshold as ordinates. For example, the zero loudness level contour corresponds to points where the curves of Figs. 2A to 2J intersect the abscissa axis. The number of db above these points is plotted as the ordinate in the loudness level contours shown in Fig. 3. From a consideration of the nature of the hearing mechanism we believe that these curves should be smooth. These curves, therefore,

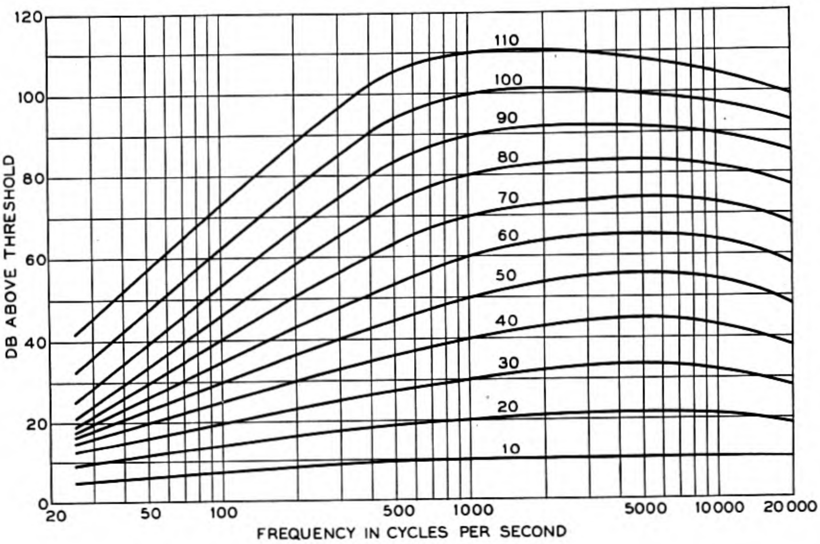


Fig. 3—Loudness level contours.

represent the best set of smooth curves which we could draw through the observed points. After the smoothing process, the curves in Figs. 2A to 2J were then adjusted to correspond. The curves shown in these figures are such adjusted curves.

In Fig. 4 a similar set of loudness level contours is shown using intensity levels as ordinates. There are good reasons [5] for believing that the peculiar shape of these contours for frequencies above 1000 c.p.s. is due to diffraction around the head of the observer as he faces the source of sound. It was for this reason that the smoothing process was done with the curves plotted with the level above threshold as the ordinate.

[5] Loc. cit.

From these loudness level contours, the curves shown in Figs. 5A and 5B were obtained. They show the loudness level *vs.* intensity level with frequency as a parameter. They are convenient to use for calculation purposes.

It is interesting to note that through a large part of the practical range for tones of frequencies from 300 c.p.s. to 4000 c.p.s. the loudness level is approximately equal to the intensity level. From these curves, it is possible to obtain any value of L_k in terms of β_k and f_k.

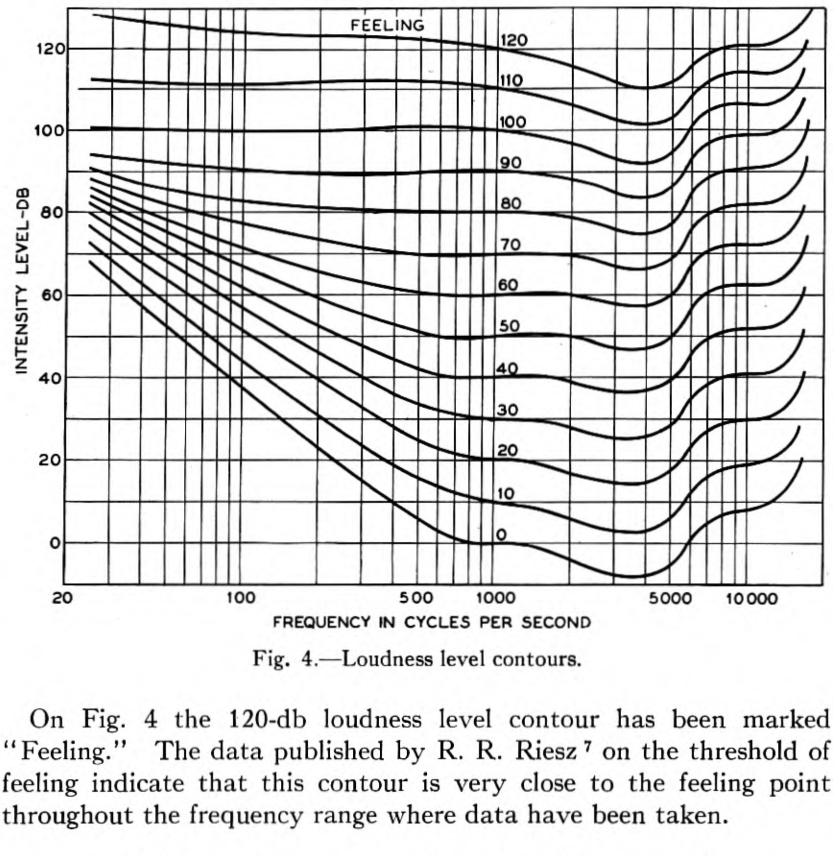


Fig. 4.—Loudness level contours.

On Fig. 4 the 120-db loudness level contour has been marked "Feeling." The data published by R. R. Riesz [7] on the threshold of feeling indicate that this contour is very close to the feeling point throughout the frequency range where data have been taken.

Determination of the Loudness Function G

In the section "Formulation of the Empirical Theory for Calculating the Loudness of a Steady Complex Tone," the fundamental equation for calculating the loudness level of a complex tone was derived,

[7] R. R. Riesz, "The Relationship Between Loudness and the Minimum Perceptible Increment of Intensity," *Jour. Acous. Soc. Am.* **4,** 211 (1933).

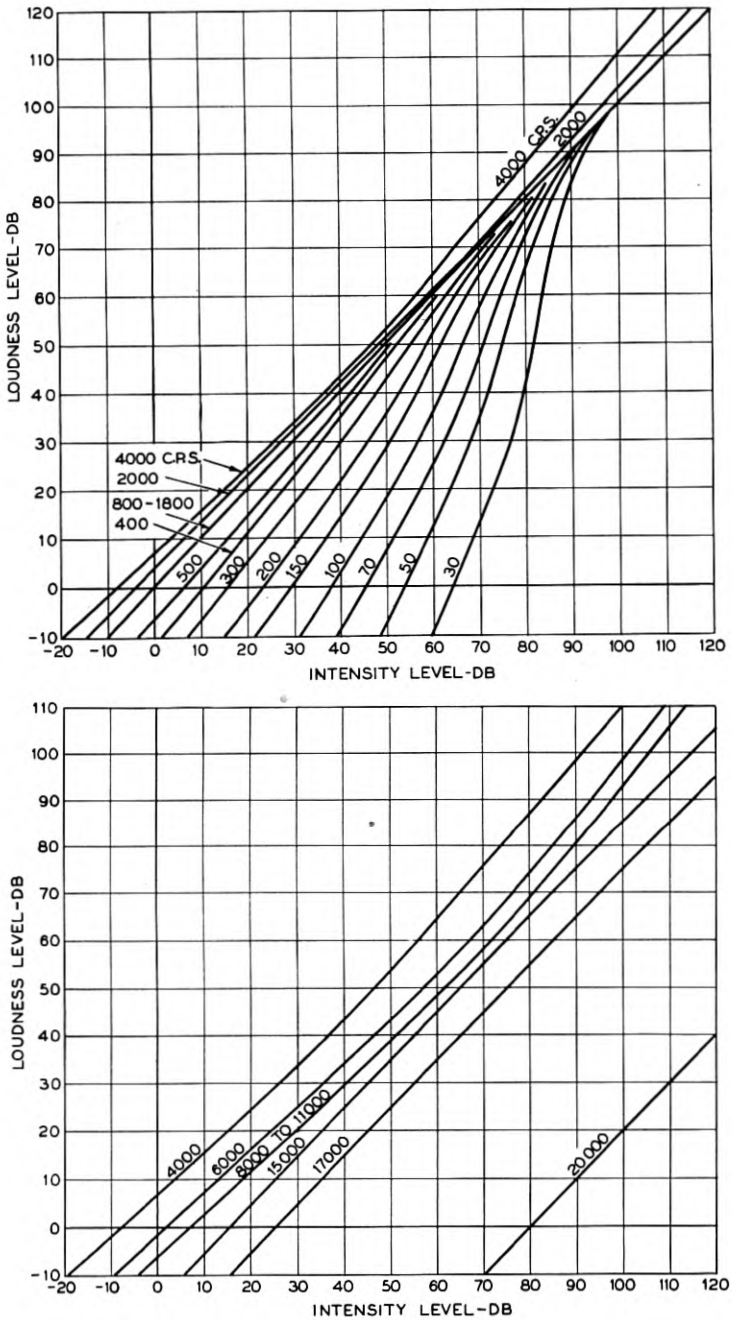


Fig. 5 (A and B)—Loudness levels of pure tones.

namely,

$$G(1000, L) = \sum_{k=1}^{k=n} b_k G(1000, L_k). \tag{10}$$

If the type of complex tone can be chosen so that b_k is unity and also so that the values of L_k for each component are equal, then the fundamental equation for calculating loudness becomes

$$G(L) = nG(L_k), \tag{14}$$

where n is the number of components. Since we are always dealing in this section with $G(1000, L)$ or $G(1000, L_k)$, the 1000 is left out in the above nomenclature. If experimental measurements of L corresponding to values of L_k are taken for a tone fulfilling the above conditions throughout the audible range, the function G can be determined. If we accept the theory that, when two simple tones widely separated in frequency act upon the ear, the nerve terminals stimulated by each are at different portions of the basilar membrane, then we would expect the interference of the loudness of one upon that of the other would be negligible. Consequently, for such a combination b is unity. Measurements were made upon two such tones, the two components being equally loud, the first having frequencies of 1000 and 2000 cycles and the second, frequencies of 125 and 1000 cycles. The observed points are shown along the second curve from the top of Fig. 6. The abscissae give the loudness level L_k of each component and the ordinates the loudness level L of the two components combined. The equation $G(y) = 2G(x)$ should represent these data. Similar measurements were made with a complex tone having 10 components, all equally loud. The method of generating such tones is described in Appendix C. The results are shown by the points along the top curve of Fig. 6. The equation $G(y) = 10G(x)$ should represent these data except at high levels where b_k is not unity.

There is probably a complete separation between stimulated patches of nerve endings when the first component is introduced into one ear and the second component into the other ear. In this case the same or different frequencies can be used. Since it is easier to make loudness balances when the same kind of sound is used, measurements were made (1) with 125-cycle tones (2) with 1000-cycle tones and (3) with 4000-cycle tones. The results are shown on Fig. 7. In this curve the ordinates give the loudness levels when one ear is used while the abscissae give the corresponding loudness levels for the same intensity level of the tone when both ears are used for listening. If binaural versus monaural loudness data actually fit into this scheme of calcula-

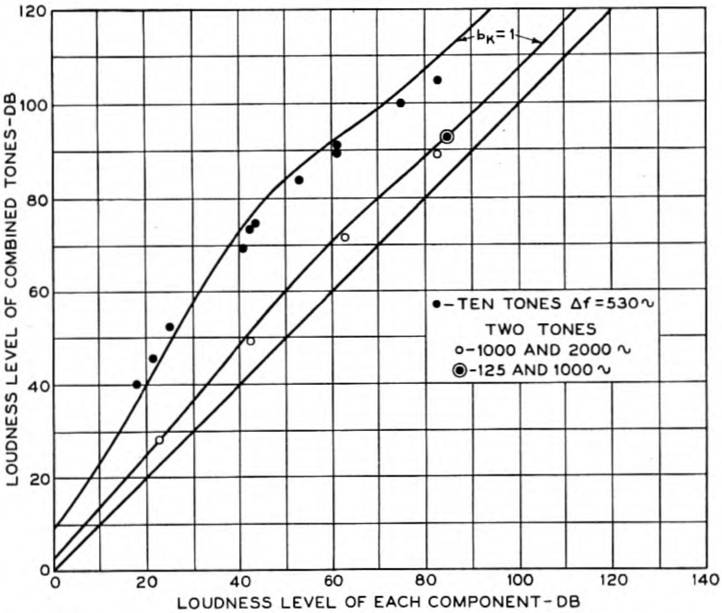


Fig. 6—Complex tones having components widely separated in frequency.

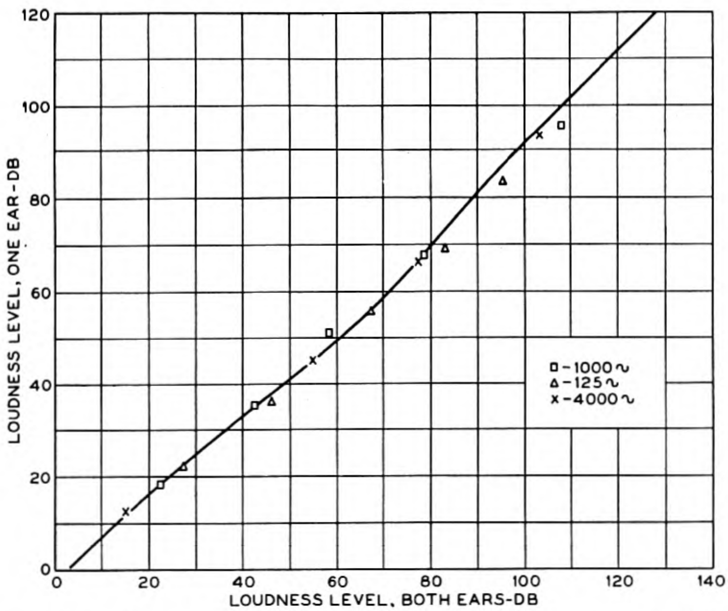


Fig. 7—Relation between loudness levels listening with one ear and with both ears.

tion these points should be represented by

$$G(y) = \tfrac{1}{2}G(x).$$

Any one of these curves which was accurately determined would be sufficient to completely determine the function G.

For example, consider the curve for two tones. It is evident that it is only necessary to deal with relative values of G so that we can choose one value arbitrarily. The value of $G(0)$ was chosen equal to unity. Therefore,

$$\begin{aligned}
&G(0) = 1, \\
&G(y_0) = 2G(0) = 2 && \text{where } y_0 \text{ corresponds to } x = 0, \\
&G(y_1) = 2G(x_1) = 2G(y_0) = 4 && \text{where } y_1 \text{ corresponds to } x_1 = y_0, \\
&G(y_2) = 2G(x_2) = 2G(y_1) = 8 && \text{where } y_2 \text{ corresponds to } x_2 = y_1, \\
&G(y_k) = 2G(x_k) = 2G(y_{k-1}) = 2^{k+1} && \text{where } y_k \text{ corresponds to } x_k = y_{k-1}.
\end{aligned}$$

In this way a set of values for G can be obtained. A smooth curve connecting all such calculated points will enable one to find any value of $G(x)$ for a given value of x. In a similar way sets of values can be obtained from the other two experimental curves. Instead of using any one of the curves alone the values of G were chosen to best fit all three sets of data, taking into account the fact that the observed points for the 10-tone data might be low at the higher levels where b would be less than unity. The values for the function which were finally adopted are given in Table III. From these values the three solid curves of Figs. 6 and 7 were calculated by the equations

$$G(y) = 10G(x), \qquad G(y) = 2G(x), \qquad G(y) = \tfrac{1}{2}G(x).$$

The fit of the three sets of data is sufficiently good, we think, to justify the point of view taken in developing the formula. The calculated points for the 10-component tones agree with the observed ones when the proper value of b_k is introduced into the formula. In this connection it is important to emphasize that in calculating the loudness level of a complex tone under the condition of listening with one ear instead of two, a factor of $\frac{1}{2}$ must be placed in front of the summation of Eq. (10). This will be explained in greater detail later. The values of G for negative values of L were chosen after considering all the data on the threshold values of the complex tones studied. These data will be given with the other loudness data on complex tones. It is interesting to note here that the threshold data show that 10 pure tones, which are below the threshold when sounded separately, will combine

TABLE III

VALUES OF $G(L_k)$.

L	0	1	2	3	4	5	6	7	8	9
−10	0.015	0.025	0.04	0.06	0.09	0.14	0.22	0.32	0.45	0.70
0	1.00	1.40	1.90	2.51	3.40	4.43	5.70	7.08	9.00	11.2
10	13.9	17.2	21.4	26.6	32.6	39.3	47.5	57.5	69.5	82.5
20	97.5	113	131	151	173	197	222	252	287	324
30	360	405	455	505	555	615	675	740	810	890
40	975	1060	1155	1250	1360	1500	1640	1780	1920	2070
50	2200	2350	2510	2680	2880	3080	3310	3560	3820	4070
60	4350	4640	4950	5250	5560	5870	6240	6620	7020	7440
70	7950	8510	9130	9850	10600	11400	12400	13500	14600	15800
80	17100	18400	19800	21400	23100	25000	27200	29600	32200	35000
90	38000	41500	45000	49000	53000	57000	62000	67500	74000	81000
100	88000	97000	106000	116000	126000	138000	150000	164000	180000	197000
110	215000	235000	260000	288000	316000	346000	380000	418000	460000	506000
120	556000	609000	668000	732000	800000	875000	956000	1047000	1150000	1266000

to give a tone which can be heard. When the components are all in the high pitch range and all equally loud, each component may be from 6 to 8 db below the threshold and the combination will still be audible. When they are all in the low pitch range they may be only 2 or 3 db below the threshold. The closeness of packing of the components also influences the threshold. For example, if the ten components are all within a 100-cycle band each one may be down 10 db. It will be shown that the formula proposed above can be made to take care of these variations in the threshold.

There is still another method which might be used for determining this loudness function $G(L)$, provided one's judgment as to the magnitude of an auditory sensation can be relied upon. If a person were asked to judge when the loudness of a sound was reduced to one half it might be expected that he would base his judgment on the experience of the decrease in loudness when going from the condition of listening with both ears to that of listening with one ear. Or, if the magnitude of the sensation is the number of nerve discharges reaching the brain per second, then when this has decreased to one half, he might be able to say that the loudness has decreased one half.

In any case, if it is assumed that an observer can judge when the magnitude of the auditory sensation, that is, the loudness, is reduced to one half, then the value of the loudness function G can be computed from such measurements.

Several different research workers have made such measurements. The measurements are somewhat in conflict at the present time so that they did not in any way influence the choice of the loudness function. Rather we used the loudness function given in Table III to calculate what such observations should give. A comparison of the calculated and observed results is given below. In Table IV is shown a comparison of calculated and observed results of data taken by Ham and Parkinson.[8] The observed values were taken from Tables 1a, 1b, 2a, 2b, 3a and 3b of their paper. The calculation is very simple. From the number of decibels above threshold S the loudness level L is determined from the curves of Fig. 3. The fractional reduction is just the fractional reduction in the loudness function for the corresponding values of L. The agreement between observed and calculated results is remarkably good. However, the agreement with the data of Laird, Taylor and Wille is very poor, as is shown by Table V. The calculation was made only for the 1024-cycle tone. The observed data were taken from Table VII of the paper by Laird,

[8] L. B. Ham and J. S. Parkinson, "Loudness and Intensity Relations," *Jour. Acous. Soc. Am.* **3**, 511 (1932).

TABLE IV

COMPARISON OF CALCULATED AND OBSERVED FRACTIONAL LOUDNESS (HAM AND PARKINSON)

350 Cycles

S	L	G	Fractional Reduction in Loudness	
			Cal. %	Obs. %
74.0	85	25,000	100	100.0
70.4	82	19,800	79	83.0
67.7	79	15,800	63	67.0
64.0	75	11,400	46	49.0
59.0	70	7,950	32	35.0
54.0	65	5,870	24	26.0
44.0	53	2,680	11	15.0
34.0	41	1,100	4	8.0
59.5	71	8,510	100	100.0
57.7	69	7,440	87	92.0
55.0	66	6,240	73	77.0
49.0	59	4,070	48	57.0
44.0	53	2,680	31	38.0
39.0	47	1,780	21	25.0
34.0	41	1,060	12	13.0
24.0	29	324	4	6.0

1000 Cycles

86.0	86	27,200	100	100.0
82.4	82	19,800	73	68.0
79.7	80	17,100	63	53.0
76.0	76	12,400	46	41.0
71.0	71	8,510	31	26.0
66.0	66	6,420	24	20.0
56.0	56	3,310	12	13.0
46.0	46	1,640	6	8.0
56.0	56	3,310	100	100.0
54.2	54	2,880	87	93.4
51.5	52	2,510	76	74.6
48.8	49	2,070	62	55.0
46.0	46	1,640	49	40.9
41.0	41	1,060	32	24.5
36.0	36	675	20	10.8

2500 Cycles

74.0	69	7,440	100	100.0
70.4	64	5,560	75	86.4
67.7	62	4,950	67	68.1
64.0	58	3,820	51	49.5
59.0	53	2,680	36	32.8
54.0	48	1,920	26	23.3
44.0	39	890	12	13.0
34.0	30	360	5	6.7
44.0	39	890	100	100.0
42.2	37	740	83	94.6
39.5	36	675	76	82.2
36.8	33	505	57	61.1
34.0	30	360	41	46.0
29.0	26	222	25	27.8
24.0	21	113	13	14.9

TABLE V

COMPARISON OF CALCULATED AND OBSERVED FRACTIONAL LOUDNESS (LAIRD, TAYLOR AND WILLE)

Original Loudness Level	Level for ½ Loudness Reduction		Cal. Level for ¼ Loudness Reduction
	Cal.	Obs.	
100	92	76.0	84
90	82	68.0	73
80	71	60.0	60
70	58	49.5	48
60	50	40.5	41
50	42	31.0	34
40	33	21.0	27
30	25	14.9	20
20	16	6.5	13
10	7	5.0	4

Taylor and Wille.[9] As shown in Table V the calculation of the level for one fourth reduction in loudness agrees better with the observed data corresponding to one half reduction in loudness.

Firestone and Geiger reported some preliminary values which were in closer agreement with those obtained by Parkinson and Ham, but their completed paper has not yet been published.[10] Because of the lack of agreement of observed data of this sort we concluded that it could not be used for influencing the choice of the values of the loudness function adopted and shown in Table III. It is to be hoped that more data of this type will be taken until there is a better agreement between observed results of different observers. It should be emphasized here that changes of the level above threshold corresponding to any fixed increase or decrease in loudness will, according to the theory outlined in this paper, depend upon the frequency of the tone when using pure tones, or upon its structure when using complex tones.

DETERMINATION OF THE FORMULA FOR CALCULATING b_k

Having now determined the function G for all values of L or L_k we can proceed to find methods of calculating b_k. Its value is evidently dependent upon the frequency and intensity of all the other components present as well as upon the component being considered. For practical computations, simplifying assumptions can be made. In most cases the reduction of b_k from unity is principally due to the adjacent component on the side of the lower pitch. This is due to the fact that a tone masks another tone of higher pitch very much more

[9] Laird, Taylor and Wille, "The Apparent Reduction in Loudness," *Jour. Acous. Soc. Am.* **3**, 393 (1932).

[10] This paper is now available. P. H. Geiger and F. A. Firestone, "The Estimation of Fractional Loudness," *Jour. Acous. Soc. Am.* **5**, 25 (1933).

than one of lower pitch. For example, in most cases a tone which is 100 cycles higher than the masking tone would be masked when it is reduced 25 db below the level of the masking tone, whereas a tone 100 cycles lower in frequency will be masked only when it is reduced from 40 to 60 db below the level of the masking tone. It will therefore be assumed that the neighboring component on the side of lower pitch which causes the greatest masking will account for all the reduction in b_k. Designating this component with the subscript m, meaning the masking component, then we have b_k expressed as a function of the following variables.

$$b_k = B(f_k, f_m, S_k, S_m), \tag{15}$$

where f is the frequency and S is the level above threshold. For the case when the level of the kth component is T db below the level of the masking component, where T is just sufficient for the component to be masked, then the value of b would be equal to zero. Also, it is reasonable to assume that when the masking component is at a level somewhat less than T db below the kth component, the latter will have a value of b_k which is unity. It is thus seen that the fundamental of a series of tones will always have a value of b_k equal to unity.

For the case when the masking component and the kth component have the same loudness, the function representing b_k will be considerably simplified, particularly if it were also found to be independent of f_k and only dependent upon the difference between f_k and f_m. From the theory of hearing one would expect that this would be approximately true for the following reasons:

The distance in millimeters between the positions of maximum response on the basilar membrane for the two components is more nearly proportional to differences in pitch than to differences in frequency. However, the peaks are sharpest in the high frequency regions where the distances on the basilar membrane for a given Δf are smallest. Also, in the low frequency region where the distances for a given Δf are largest, these peaks are broadest. These two factors tend to make the interference between two components having a fixed difference in frequency approximately the same regardless of their position on the frequency scale. However, it would be extraordinary if these two factors just balanced. To test this point three complex tones having ten components with a common Δf of 50 cycles were tested for loudness. The first had frequencies of 50–100–150$\cdots$500, the second 1400–1450$\cdots$1900, and the third 3400–3450$\cdots$3900. The results of these tests are shown in Fig. 8. The abscissae give the loudness level of each component and the ordinates the measured loudness level of the combined tone. Similar results were obtained with a complex tone having ten components of equal loudness and a common frequency difference of 100 cycles. The results are shown in Fig. 9. It will be seen that although the points corresponding to the different

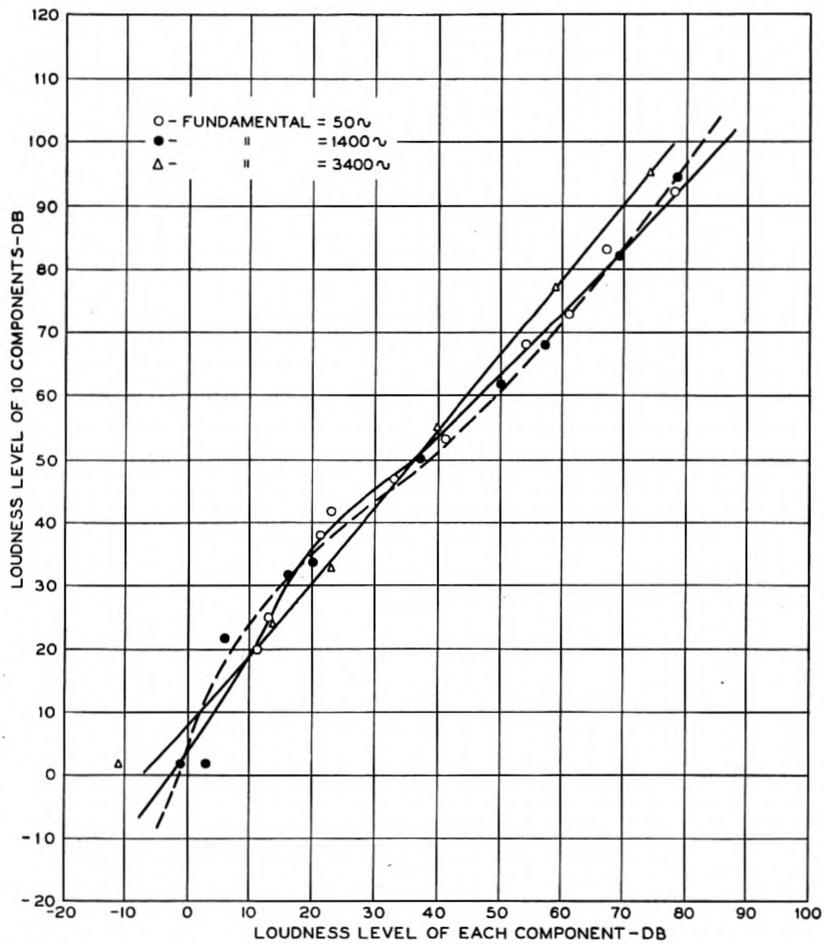


Fig. 8—Loudness levels of complex tones having ten equally loud components 50 cycles apart.

frequency ranges lie approximately upon the same curve through the middle range, there are consistent departures at both the high and low intensities. If we choose the frequency of the components largely in the middle range then this factor b will be dependent only upon Δf and L_k.

To determine the value of b for this range in terms of Δf and L_k, a series of loudness measurements was made upon complex tones having ten components with a common difference in frequency Δf and all having a common loudness level L_k. The values of Δf were 340, 230,

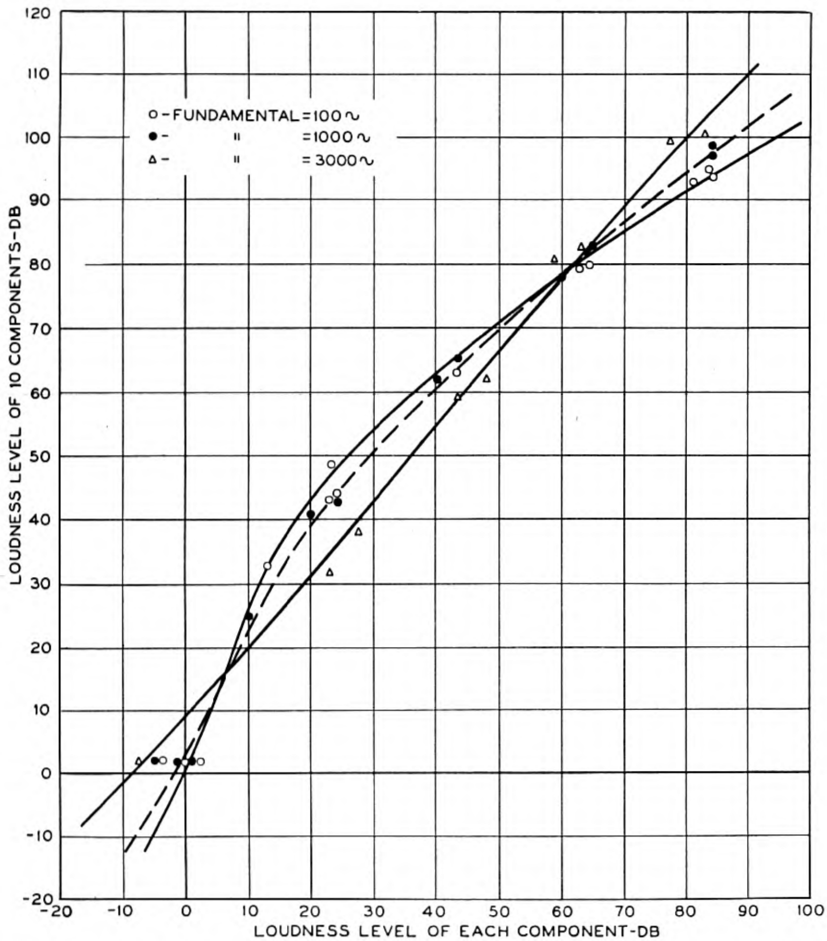


Fig. 9—Loudness levels of complex tones having ten equally loud components 100 cycles apart.

112 and 56 cycles per second. The fundamental for each tone was close to 1000 cycles. The ten-component tones having frequencies which are multiples of 530 was included in this series. The results of loudness balances are shown by the points in Fig. 10.

By taking all the data as a whole, the curves were considered to give the best fit. The values of b were calculated from these curves as follows:

According to the assumptions made above, the component of lowest pitch in the series of components always has a value of b_k equal to unity. Therefore for the series of 10 components having a common

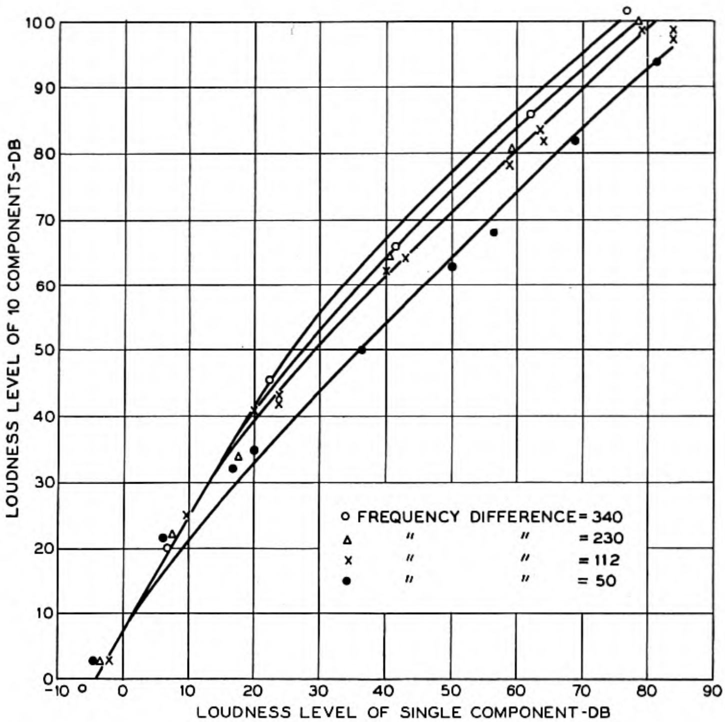


Fig. 10—Loudness levels of complex tones having ten equally loud components with a fundamental frequency of 1000 c.p.s.

loudness level L_k, the value of L is related to L_k by

$$G(L) = (1 + 9b_k)G(L_k)$$

or by solving for b_k

$$b_k = (1/9)[G(L)/G(L_k) - 1]. \qquad (16)$$

The values of b_k can be computed from this equation from the observed values of L and L_k by using the values of G given in Table III. Because of the difficulty in obtaining accurate values of L and L_k such computed values of b_k will be rather inaccurate. Consequently, considerable freedom is left in choosing a simple formula which will represent the results. When the values of b_k derived in this way were plotted with b_k as ordinates and Δf as abscissae and L_k as a variable parameter then the resulting graphs were a series of straight lines going through the common point $(-250, 0)$ but having slopes depending upon L_k. Consequently the following formula

$$b_k = [(250 + \Delta f)/1000]Q(L_k) \tag{17}$$

will represent the results. The quantity Δf is the common difference in frequency between the components, L_k the loudness level of each component, and Q a function depending upon L_k. The results indicated that Q could be represented by the curve in Fig. 11.

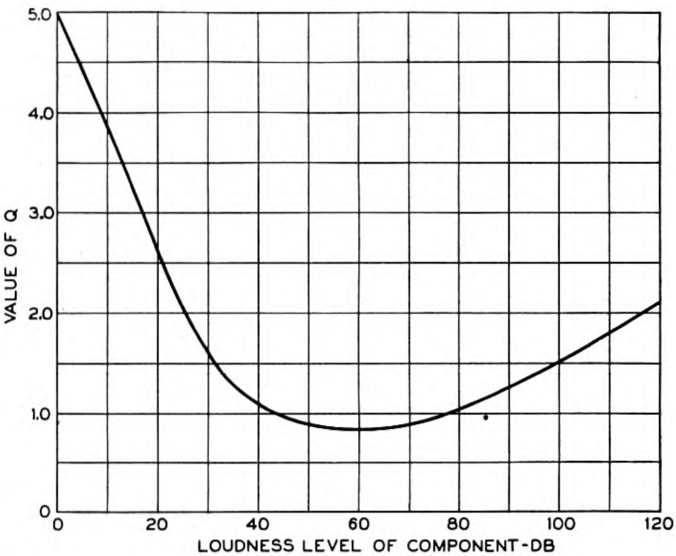


Fig. 11—Loudness factor Q.

Also the condition must be imposed upon this equation that b is always taken as unity whenever the calculation gives values greater than unity. The solid curves shown in Fig. 10 are actually calculated curves using these equations, so the comparison of these curves with the observed points gives an indication of how well this equation fits the data. For this series of tones Q could be made to depend upon β_k rather than L_k and approximately the same results would be obtained since β_k and L_k are nearly equal in this range of frequencies. However, for tones having low intensities and low frequencies, β_k will be much larger than L_k and consequently Q will be smaller and hence the calculated loudness smaller. The results in Figs. 8 and 9 are just contrary to this. To make the calculated and observed results agree with these two sets of data, Q was made to depend upon

$$x = \beta + 30 \log f - 95$$

instead of L_k.

It was found when using this function of β and f as an abscissa and the same ordinates as in Fig. 10, a value of Q was obtained which gives just as good a fit for the data of Fig. 10 and also gives a better fit for the data of Figs. 8 and 9. Other much more complicated factors were tried to make the observed and calculated results shown in these two figures come into better agreement but none were more satisfactory than the simple procedure outlined above. For purpose of calculation the values of Q are tabulated in Table VI.

TABLE VI

Values of $Q(X)$

X	0	1	2	3	4	5	6	7	8	9
0	5.00	4.88	4.76	4.64	4.53	4.41	4.29	4.17	4.05	3.94
10	3.82	3.70	3.58	3.46	3.35	3.33	3.11	2.99	2.87	2.76
20	2.64	2.52	2.40	2.28	2.16	2.05	1.95	1.85	1.76	1.68
30	1.60	1.53	1.47	1.40	1.35	1.30	1.25	1.20	1.16	1.13
40	1.09	1.06	1.03	1.01	0.99	0.97	0.95	0.94	0.92	0.91
50	0.90	0.90	0.89	0.89	0.88	0.88	0.88	0.88	0.88	0.88
60	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.89	0.89	0.90
70	0.90	0.91	0.92	0.93	0.94	0.96	0.97	0.99	1.00	1.02
80	1.04	1.06	1.08	1.10	1.13	1.15	1.17	1.19	1.22	1.24
90	1.27	1.29	1.31	1.34	1.36	1.39	1.41	1.44	1.46	1.48
100	1.51	1.53	1.55	1.58	1.60	1.62	1.64	1.67	1.69	1.71

Note: $X = \beta_k + 30 \log f_k - 95$.

There are reasons based upon the mechanics of hearing for treating components which are very close together by a separate method. When they are close together the combination must act as though the energy were all in a single component, since the components act upon approximately the same set of nerve terminals. For this reason it seems logical to combine them by the energy law and treat the combination as a single frequency. That some such procedure is necessary is shown from the absurdities into which one is led when one tries to make Eq. (17) applicable to all cases. For example, if 100 components were crowded into a 1000-cycle space about a 1000-cycle tone, then it is obvious that the combination should sound about 20 db louder. But according to Eq. (10) to make this true for values of L_k greater than 45, b_k must be chosen as 0.036. Similarly, for 10 tones thus crowded together $L - L_k$ must be about 10 db and therefore $b_k = 0.13$ and then for two such tones $L - L_k$ must be 3 db and the corresponding value of $b_k = 0.26$. These three values must belong to the same condition $\Delta f = 10$. It is evident then that the formulae for b given by Eq. (17) will lead to very erroneous results for such components.

In order to cover such cases it was necessary to group together all components within a certain frequency band and treat them as a single component. Since there was no definite criterion for determining accurately what these limiting bands should be, several were tried and ones selected which gave the best agreement between computed and observed results. The following band widths were finally chosen:

For frequencies below 2000 cycles, the band width is 100 cycles; for frequencies between 2000 and 4000 cycles, the band width is 200 cycles; for frequencies between 4000 and 8000 cycles, the band width is 400 cycles; and for frequencies between 8000 and 16,000 cycles, the band width is 800 cycles. If there are k components within one of these limiting bands, the intensity I taken for the equivalent single frequency component is given by

$$I = \sum I_k = \sum 10^{\beta_k/10}. \tag{18}$$

A frequency must be assigned to the combination. It seems reasonable to assign a weighted value of f given by the equation

$$f = \sum f_k I_k / I = \sum f_k 10^{\beta_k/10} / \sum 10^{\beta/10}. \tag{19}$$

Only a small error will be introduced if the mid-frequency of such bands be taken as the frequency of an equivalent component except for the band of lowest frequency. Below 125 cycles it is important that the frequency and intensity of each component be known, since in this region the loudness level L_k changes very rapidly with both changes in intensity and frequency. However, if the intensity for this band is lower than that for other bands, it will contribute little to the total loudness so that only a small error will be introduced by a wrong choice of frequency for the band.

This then gives a method of calculating b_k when the adjacent components are equal in loudness. When they are not equal let us define the difference ΔL by

$$\Delta L = L_k - L_m. \tag{20}$$

Also let this difference be T when L_m is adjusted so that the masking component just masks the component k. Then the function for calculating b must satisfy the following conditions:

$$b_k = [(250 + \Delta f)/1000]Q \qquad \text{when } \Delta L = 0,$$

$$b_k = 0 \qquad \text{when } \Delta L = -T.$$

Also the following condition when L_k is larger than L_m must be satisfied, namely, $b_k = 1$ when ΔL = some value somewhat smaller than $+ T$. The value of T can be obtained from masking curves. An examination of these data indicates that to a good approximation the value of T is dependent upon the single variable $f_k - 2f_m$. A curve showing the relation between T and this variable is shown in Fig. 12. It will be seen that for most practical cases the value of T is 25. It cannot be claimed that the curve of Fig. 12 is an accurate representation of the masking data, but it is sufficiently accurate for the purpose of loudness calculation since rather large changes in T will produce a very slight change in the final calculated loudness level.

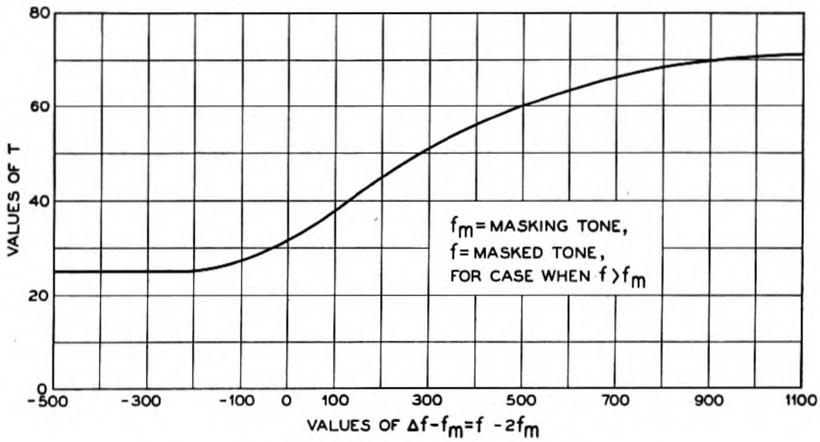


Fig. 12—Values of the masking T.

Data were taken in an effort to determine how this function depended upon ΔL but it was not possible to obtain sufficient accuracy in the experimental results. The difference between the resultant loudness level when half the tones are down so as not to contribute to loudness and when these are equal is not more than 4 or 5 db, which is not much more than the observational errors in such results.

A series of tests were made with tones similar to those used to obtain the results shown in Figs. 8 and 9 except that every other component was down in loudness level 5 db. Also a second series was made in which every other component was down 10 db. Although these data were not used in determining the function described above, it was useful as a check on the final equations derived for calculating the loudness of tones of this sort.

The factor finally chosen for representing the dependence of b_k upon ΔL is $10^{\Delta L/T}$. This factor is unity for $\Delta L = 0$, fulfilling the first condition mentioned above. It is 0.10 instead of zero for $\Delta L = -25$, the most probable value of T. For $\Delta f = 100$ and $Q = 0.88$ we will obtain the smallest value of b_k without applying the ΔL factor, namely, 0.31. Then when using this factor as given above, all values of b_k will be unity for values of ΔL greater than 12 db.

Several more complicated functions of ΔL were tried but none of them gave results showing a better agreement with the experimental values than the function chosen above.

The formula for calculation of b_k then becomes

$$b_k = [(250 + f_k - f_m)/1000]10^{(L_k - L_m)/T}Q(\beta_k + 30 \log f_k - 95) \quad (21)$$

where

f_k is the frequency of the component expressed in cycles per second,

f_m is the frequency of the masking component expressed in cycles per second,

L_k is the loudness level of the kth component when sounding alone,

L_m is the loudness level of the masking tone,

Q is a function depending upon the intensity level β_k and the frequency f_k of each component and is given in Table VI as a function of $x = \beta_k + 30 \log f_k - 95$,

T is the masking and is given by the curve of Fig. 12.

It is important to remember *that b_k can never be greater than unity so that all calculated values greater than this must be replaced with values equal to unity.* Also all components *within the limiting frequency bands must be grouped together as indicated above.* It is very helpful to remember that any component for which the loudness level is 12 db below the kth component, that is, the one for which b is being calculated, need not be considered as possibly being the masking component. If all the components preceding the kth are in this class then b_k is unity.

Recapitulation

With these limitations the formula for calculating the loudness level L of a steady complex tone having n components is

$$G(L) = \sum_{k=1}^{k=n} b_k G(L_k), \quad (10)$$

where b_k is given by Eq. (21). If the values of f_k and β_k are measured directly then corresponding values of L_k can be found from Fig. 5.

Having these values, the masking component can be found either by inspection or better by trial in Eq. (21). That component whose values of L_m, f_m and T introduced into this equation gives the smallest value of b_k is the masking component.

The values of G and Q can be found from Tables III and VI from the corresponding values of L_k, β_k, and f_k. If all these values are now introduced into Eq. (10), the resulting value of the summation is the *loudness* of the complex tone. The loudness level L corresponding to it is found from Table III.

If it is desired to know the loudness obtained if the typical listener used only one ear, the result will be obtained if the summation indicated in Eq. (10) is divided by 2. Practically the same result will be obtained in most instances if the loudness level L_k for each component when listened to with one ear instead of both ears is inserted in Eq. (10). ($G(L_k)$ for one ear listening is equal to one half $G(L_k)$ for listening with both ears for the same value of the intensity level of the component.) If two complex tones are listened to, one in one ear and one in the other, it would be expected that the combined loudness would be the sum of the two loudness values calculated for each ear as though no sound were in the opposite ear, although this has not been confirmed by experimental trial. In fact, the loudness reduction factor b_k has been derived from data taken with both ears only, so strictly speaking, its use is limited to this type of listening.

To illustrate the method of using the formula the loudness of two complex tones will be calculated. The first may represent the hum from a dynamo. Its components are given in the table of computations.

COMPUTATIONS

k	f_k	β_k	L_k	G_k	b_k	
1	60	50	3	3	1.0	
2	180	45	25	197	1.0	$\Sigma b_k G_k = 1009$
3	300	40	30	360	1.0	
4	540	30	27	252	1.0	$L = 40$
5	1200	25	25	197	1.0	

The first step is to find from Fig. 5 the values of L_k from f_k and β_k. Then the loudness values G_k are found from Table III. Since the values of L are low and the frequency separation fairly large, one familiar with these functions would readily see that the values of b would be unity and a computation would verify it so that the sum of the G values gives the total loudness 1009. This corresponds to a loudness level of 40.

The second tone calculated is this same hum amplified 30 db. It better illustrates the use of the formula.

COMPUTATIONS

k	f_k	β_k	L_k	G_k	f_m	L_m	$(30 \log f_k - 95)$	Q	b	$b \times G$
1	60	80	69	7440	—	—	—	—	1.00	7440
2	180	75	72	9130	60	69	−28	0.91	0.41	3740
3	300	70	69	7440	180	72	−21	0.91	0.27	2010
4	540	60	60	4350	300	69	−13	0.94	0.23	1000
5	1200	55	55	3080	540	60	− 3	0.89	0.61	1880
									loudness G =	16070
									loudness level L =	79 db

The loudness level of the combined tones is only 7 db above the loudness level of the second component. If only one ear is used in listening, the loudness of this tone is one half, corresponding to a loudness level of 70 db.

Comparison of Observed and Calculated Results on the Loudness Levels of Complex Tones

In order to show the agreement between observed loudness levels and levels calculated by means of the formula developed in the preceding sections, the results of a large number of tests are given here, including those from which the formula was derived. In Tables VII to XIII, the first column shows the frequency range over which the components of the tones were distributed, the figures being the frequencies of the first and last components. Several tones having two components were tested, but as the tables indicate, the majority of the tones had ten components. Because of a misunderstanding in the

TABLE VII

Two Component Tones ($\Delta L = 0$)

Frequency Range	Δf	Loudness Levels (db)					
1000–1100	100	L_k	83	63	43	23	2
		$L_{obs.}$	87	68	47	28	2
		$L_{calc.}$	87	68	47	28	4
1000–2000	1000	L_k	83	63	43	23	−1
		$L_{obs.}$	89	71	49	28	2
		$L_{calc.}$	91	74	52	28	1
125–1000	875	L_k	84				
		$L_{obs.}$	92				
		$L_{calc.}$	92				

TABLE VIII

TEN COMPONENT TONES ($\Delta L = 0$)

Frequency Range	Δf	Loudness Levels (db)										
		L_k	67	54	33	21	11	−1				
50–500	50	$L_{obs.}$	83	68	47	38	20	2				
		$L_{calc.}$	81	72	53	39	24	8				
		L_k	78	61	41	23	13	−1				
50–500	50	$L_{obs.}$	92	73	53	42	25	2				
		$L_{calc.}$	91	77	60	42	27	8				
		L_k	78	69	50	16	6	−1				
1400–1895	55	$L_{obs.}$	94	82	62	32	22	2				
		$L_{calc.}$	93	83	65	31	17	0				
		L_k	57	37	20	3						
1400–1895	55	$L_{obs.}$	68	50	34	2						
		$L_{calc.}$	73	52	36	5						
		L_k	84	64	43	24	2	84	64	43	24	2
100–1000	100	$L_{obs.}$	95	83	59	41	2	94	80	63	44	2
		$L_{calc.}$	100	83	68	47	12	100	83	68	47	12
		L_k	81	64	43	23	13	−4				
100–1000	100	$L_{obs.}$	93	82	65	49	33	2				
		$L_{calc.}$	98	83	68	45	27	3				
		L_k	83	63	43	23	0					
100–1000	100	$L_{obs.}$	95	79	59	43	2					
		$L_{calc.}$	99	82	68	45	9					
		L_k	83	63	43	23	78	59	48	27	−7	
3100–3900	100	$L_{obs.}$	100	82	59	32	99	81	62	38	2	
		$L_{calc.}$	100	80	60	38	95	77	65	42	0	
		L_k	79	60	41	17	7	−4				
1100–3170	230	$L_{obs.}$	100	81	65	33	22	2				
		$L_{calc.}$	100	83	64	34	18	3				
		L_k	79	62	42	23	13	−2				
260–2600	260	$L_{obs.}$	97	82	65	44	28	2				
		$L_{calc.}$	100	85	68	45	27	5				
		L_k	75	53	43	25	82	61	43	17	−2	
530–5300	530	$L_{obs.}$	100	83	73	52	105	90	73	40	2	
		$L_{calc.}$	101	82	72	48	108	89	72	34	5	
		L_k	61	41	21	−3						
530–5300	530	$L_{obs.}$	89	69	45	2						
		$L_{calc.}$	89	70	42	4						

design of the apparatus for generating the latter tones, a number of them contained eleven components, so for purposes of identification, these are placed in a separate group. In the second column of the tables, next to the frequency range of the tones, the frequency difference (Δf) between adjacent components is given. The remainder of

TABLE IX

ELEVEN COMPONENT TONES ($\Delta L = 0$)

Frequency Range	Δf	Loudness Levels (db)						
1000–2000	100	L_k $L_{obs.}$ $L_{calc.}$	84 97 103	64 83 84	43 65 64	24 43 45	−1 2 7	
1000–2000	100	L_k $L_{obs.}$ $L_{calc.}$	84 99 103	64 82 84	43 65 64	24 42 45	1 2 11	
1150–2270	112	L_k $L_{obs.}$ $L_{calc.}$	79 99 98	60 78 81	40 62 61	20 41 40	10 25 23	−5 2 1
1120–4520	340	L_k $L_{obs.}$ $L_{calc.}$	77 102 101	62 86 88	42 66 69	22 46 44	7 20 19	−7 2 −1

the data pertains to the loudness levels of the tones. Opposite L_k are given the common loudness levels to which all the components of the tone were adjusted for a particular test, and in the next line the results of the test, that is, the observed loudness levels ($L_{obs.}$), are given. Directly beneath each observed value, the calculated loudness levels ($L_{calc.}$) are shown. The three associated values of L_k, $L_{obs.}$, and $L_{calc.}$ in each column represent the data for one complete test. For example, in Table VIII, the first tone is described as having ten components, and for the first test shown each component was adjusted to have a loudness level (L_k) of 67 db. The results of the test gave an observed loudness level ($L_{obs.}$) of 83 db for the ten components acting together, and the calculated loudness level ($L_{calc.}$) of this same tone was 81 db. The probable error of the observed results in the tables is approximately $\pm$ 2 db.

TABLE X

TEN COMPONENT TONES (ΔL = 5 db)

Frequency Range	Δf	Loudness Levels (db)						
1725–2220	55	L_k $L_{obs.}$ $L_{calc.}$	82 101 95	62 73 76	43 54 56	27 38 40	17 30 30	−6 2 −1
1725–2220	55	L_k $L_{obs.}$ $L_{calc.}$	80 94 93	62 66 76	42 50 54	22 33 35	12 22 22	−2 2 4

In the next series of data, adjacent components had a difference in loudness level of 5 db, that is, the first, third, fifth, etc., components had the loudness level given opposite L_k, and the even numbered components were 5 db lower. (Tables X and XI.)

TABLE XI

ELEVEN COMPONENT TONES (ΔL = 5 db).

Frequency Range	Δf	Loudness Levels (db)						
57–627	57	L_k	79	61	41	26	16	1
		$L_{obs.}$	91	73	56	41	28	2
		$L_{calc.}$	90	76	59	43	28	8
3420–4020	60	L_k	76	61	42	25	15	−9
		$L_{obs.}$	95	77	55	33	25	2
		$L_{calc.}$	89	75	54	36	26	−4

In the following set of tests (Tables XII and XIII) the difference in loudness level of adjacent components was 10 db.

TABLE XII

TEN COMPONENT TONES (ΔL = 10 db)

Frequency Range	Δf	Loudness Levels (db)						
1725–2220	55	L_k	79	59	40	19	9	−5
		$L_{obs.}$	95	71	54	33	22	2
		$L_{calc.}$	91	73	51	31	17	−1
1725–2220	55	L_k	79	61	41	27	17	−1
		$L_{obs.}$	89	67	48	37	27	2
		$L_{calc.}$	92	75	53	39	28	4

TABLE XIII

ELEVEN COMPONENT TONES (ΔL = 10 db)

Frequency Range	Δf	Loudness Levels (db)						
57–627	57	L_k	80	62	42	27	17	2
		$L_{obs.}$	88	70	53	40	27	2
		$L_{calc.}$	90	76	59	45	30	8
3420–4020	60	L_k	81	62	42	27	17	−4
		$L_{obs.}$	100	70	50	33	26	2
		$L_{calc.}$	94	75	53	37	27	0

The next data are the results of tests made on the complex tone generated by the Western Electric No. 3A audiometer. When analyzed, this tone was found to have the voltage level spectrum shown in Table XIV. When the r.m.s. voltage across the receivers used was unity, that is, zero voltage level, then the separate components had the voltage levels given in this table. Adding to the voltage levels the calibration constant for the receivers used in making the loudness tests gives the values of β for zero voltage level across the receivers. The values of β for any other voltage level are obtained by addition of the level desired.

TABLE XIV

Voltage Level Spectrum of No. 3A Audiometer Tone

Frequency	Voltage Level	Frequency	Voltage Level
152	− 2.1	2128	−11.4
304	− 5.4	2280	−16.9
456	− 4.7	2432	−14.1
608	− 5.9	2584	−16.2
760	− 4.6	2736	−17.4
912	− 6.8	2880	−17.5
1064	− 6.0	3040	−20.0
1216	− 8.1	3192	−19.4
1368	− 7.6	3344	−22.7
1520	− 9.1	3496	−23.7
1672	−10.0	3648	−25.6
1824	− 9.9	3800	−24.6
1976	−14.1	3952	−26.8

Tests were made on the audiometer tone with the same receivers [11] that were used with the other complex tones, but in addition, data were available on tests made about six years ago using a different type of receiver. This latter type of receiver was recalibrated (Fig. 13) and computations made for both the old and new tests. In the older set of data, levels above threshold were given instead of voltage levels, so in utilizing it here, it was necessary to assume that the threshold levels of the new and old tests were the same.

Computations were made at the levels tested experimentally and a comparison of observed and calculated results is shown in Table XV.

The agreement of observed and calculated results is poor for some of the tests, but the close agreement in the recent data at low levels and in the previous data at high levels indicates that the observed results are not as accurate as could be desired. Because of the labor involved these tests have not been repeated.

At the time the tests were made several years ago on the No. 3A Audiometer tone, the reduction in loudness level which takes place when certain components are eliminated was also determined. As this

[11] See Calibration shown in Fig. 1.

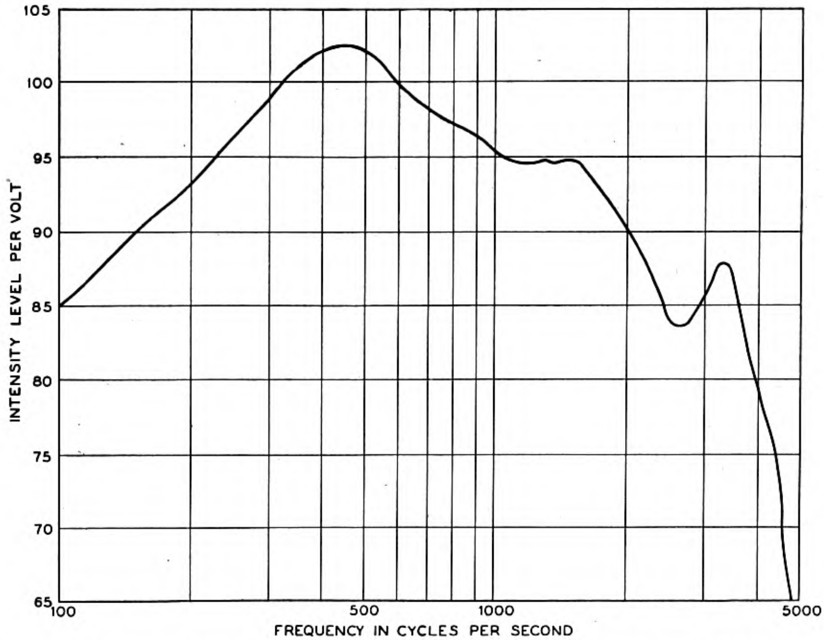


Fig. 13—Calibration of receivers for tests on the No. 3A audiometer tone

can be readily calculated with the formula developed here, a comparison of observed and calculated results will be shown. In Fig. 14A, the ordinate is the reduction in loudness level resulting when a No. 3A Audiometer tone having a loudness level of 42 db was changed by the insertion of a filter which eliminated all of the components above or below the frequency indicated on the abscissa. The observed data are the plotted points and the smooth curves are calculated results. A similar comparison is shown in Figs. 14B, C and D for other levels.

TABLE XV

A. Recent Tests on No. 3A Audiometer Tone

R.m.s. Volt. Level....	−38	−55	−59	−70	−75	−78	−80	−87	−89	−100	−102
L_{obs}............	95	85	79	61	56	41	42	28	22	2	2
L_{calc}............	89	74	71	57	49	44	40	28	25	7	4

B. Previous Tests on No. 3A Audiometer Tone

R.m.s. Volt. Level....	+10	−9	−40	−49	−60	−69	−91
L_{obs}............	118	103	77	69	61	50	2
L_{calc}............	119	103	82	73	56	41	6

This completes the data which are available on steady complex tones. It is to be hoped that others will find the field of sufficient importance to warrant obtaining additional data for improving and testing the method of measuring and calculating loudness levels.

In view of the complex nature of the problem this computation method cannot be considered fully developed in all its details and as more accurate data accumulates it may be necessary to change the formula for b. Also at the higher levels some attention must be given to phase differences between the components. However, we feel that the form of the equation is fundamentally correct and the loudness

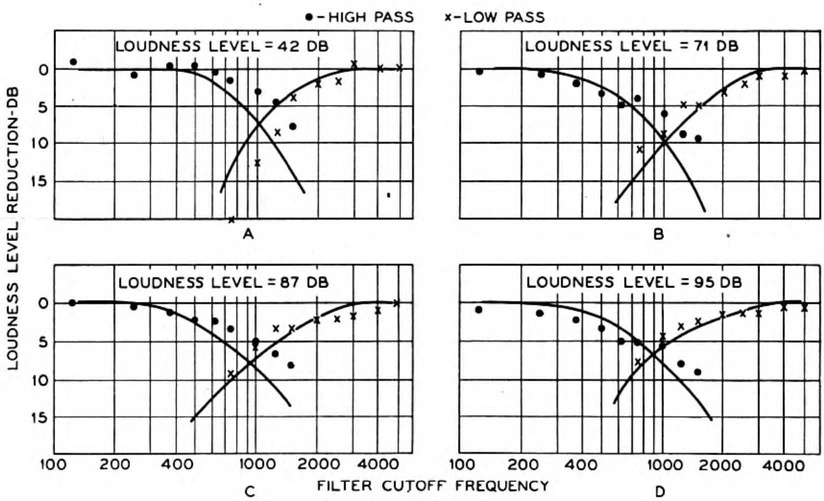


Fig. 14 (A to D)—Loudness level reduction tests on the No. 3A audiometer tone.

function, G, corresponds to something real in the mechanism of hearing. The present values given for G may be modified slightly, but we think that they will not be radically changed.

A study of the loudness of complex sounds which are not steady, such as speech and sounds of varying duration, is in progress at the present time and the results will be reported in a second paper on this subject.

Appendix A. Experimental Method of Measuring the Loudness Level of a Steady Sound

A measurement of the loudness level of a sound consists of listening alternately to the sound and to the 1000-cycle reference tone and adjusting the latter until the two are equally loud. If the intensity level of the reference tone is L decibels when this condition is reached, the sound is said to have a loudness level of L decibels. When the character of the sound being measured differs only slightly from that of the reference tone, the comparison is easily and quickly made, but for other sounds the numerous factors which enter into a judgment of equality of loudness become important, and an experimental method should be used which will yield results typical of the average normal ear and normal physiological and psychological conditions.

A variety of methods have been proposed to accomplish this, differing not only in general classification, that is, the method of average error, constant stimuli, etc., but also in important experimental details such as the control of noise conditions and fatigue effects. In some instances unique devices have been used to facilitate a ready comparison of sounds. One of these, the alternation phonometer,[12] introduces into the comparison important factors such as the duration time of the sounds and the effect of transient conditions. The merits of a particular method will depend upon the circumstances under which it is to be used. The one to be described here was developed for an extensive series of laboratory tests.

To determine when two sounds are equally loud it is necessary to rely upon the judgment of an observer, and this involves of course, not only the ear mechanism, but also associated mental processes, and effectively imbeds the problem in a variety of psychological factors. These difficulties are enhanced by the large variations found in the judgments of different observers, necessitating an investigation conducted on a statistical basis. The method of constant stimuli, wherein the observer listens to fixed levels of the two sounds and estimates which sound is the louder, seemed best adapted to control the many factors involved, when using several observers simultaneously. By means of this method, an observer's part in the test can be readily limited to an indication of his loudness judgment. This is essential as it was found that manipulation of apparatus controls, even though they were not calibrated, or participation in any way other than as a judge of loudness values, introduced undesirable factors which were aggravated by continued use of the same observers over a long period of time. Control of fatigue, memory effects, and the association of an observer's judgments with the results of the tests or with the judgments of other observers could be rigidly maintained with this method, as will be seen from the detailed explanation of the experimental procedure.

[12] D. Mackenzie, "Relative Sensitivity of the Ear at Different Levels of Loudness," *Phys. Rev.* **20,** 331 (1922).

The circuit shown in Fig. 15 was employed to generate and control the reference tone and the sounds to be measured. Vacuum tube oscillators were used to generate pure tones, and for complex tones and other sounds, suitable sources were substituted. By means of the voltage measuring circuit and the attenuator, the voltage level (voltage level = 20 log V) impressed upon the terminals of the receivers, could be determined. For example, the attenuator, which was calibrated in decibels, was set so that the voltage measuring set indicated 1 volt was being impressed upon the receiver. Then the difference between this setting and any other setting is the voltage level. To obtain the intensity level of the sound we must know the calibration of the receivers.

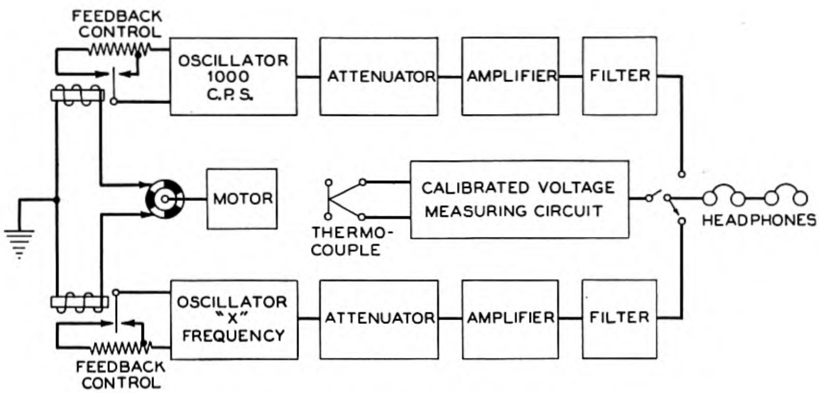


Fig. 15—Circuit for loudness balances.

The observers were seated in a sound-proof booth and were required only to listen and then operate a simple switch. These switches were provided at each position and were arranged so that the operations of one observer could not be seen by another. This was necessary to prevent the judgments of one observer from influencing those of another observer. First they heard the sound being tested, and immediately afterwards the reference tone, each for a period of one second. After a pause of one second this sequence was repeated, and then they were required to estimate whether the reference tone was louder or softer than the other sound and indicate their opinions by operating the switches. The levels were then changed and the procedure repeated. The results of the tests were recorded outside the booth.

The typical recording chart shown in Fig. 16 contains the results of three observers testing a 125-cycle tone at three different levels. Two

125 c.p.s. Pure Tone Test No. 4 Crew No. 1. 1000 c.p.s. Voltage Level (db)

Obs.		+6	+2	−2	−6	−10	−14	−18	−22	−26
125	CK	+	+	+	+	+	0	0	0	0
c.p.s.	AS	+	+	+	+	0	0	0	0	0
Volt.	DH	+	+	0	0	0	0	0	0	0
level =	CK	+	+	+	+	+	0	0	0	0
+ 9.8 db	AS	+	+	+	+	0	0	0	0	0
	DH	+	+	0	0	+	0	0	0	0
	CK	+	+	+	+	0	0	0	0	0
	AS	+	+	+	0	0	0	0	0	0
	DH	+	+	0	0	0	0	0	0	0
		0	−4	−8	−12	−16	−20	−24	−28	−32
125	CK	+	+	+	+	0	+	+	0	0
c.p.s.	AS	+	+	+	+	+	0	0	0	0
Volt.	DH	+	+	+	+	0	0	0	0	0
level =	CK	+	+	+	+	+	+	+	0	0
− 3.2 db	AS	+	+	+	+	+	+	0	0	0
	DH	+	+	+	0	+	0	+	0	0
	CK	+	+	+	+	+	+	0	0	0
	AS	+	+	+	+	+	0	0	0	0
	DH	+	+	+	0	+	0	0	0	0
		−15	−19	−23	−27	−31	−35	−39	−43	−47
125	CK	+	+	+	+	+	0	0	0	0
c.p.s.	AS	+	+	+	+	0	0	0	0	0
Volt.	DH	+	+	0	+	0	+	0	0	0
level =	CK	0	+	+	+	+	+	0	0	0
− 14.2	AS	+	+	+	+	0	+	0	0	0
db	DH	+	+	0	+	0	0	+	0	0
	CK	+	+	0	+	+	+	0	0	0
	AS	+	+	0	0	+	+	0	0	0
	DH	+	+	0	0	0	0	+	0	0

Fig. 16—Loudness balance data sheet.

marks were used for recording the observers' judgments, a cipher indicating the 125-cycle tone to be the louder, and a plus sign denoting the reference tone to be the louder of the two. No equal judgments were permitted. The figures at the head of each column give the voltage level of the reference tone impressed upon the receivers, that is, the number of decibels from 1 volt, plus if above and minus if below, and those at the side are similar values for the tone being tested. Successive tests were chosen at random from the twenty-seven possible combinations of levels shown, thus reducing the possibility of memory effects. The levels were selected so the observers listened to reference tones which were louder and softer than the tone being tested and the median of their judgments was taken as the point of equal loudness.

The data on this recording chart, combined with a similar number of observations by the rest of the crew, (a total of eleven observers) are shown in graphical form in Fig. 17. The arrow indicates the median

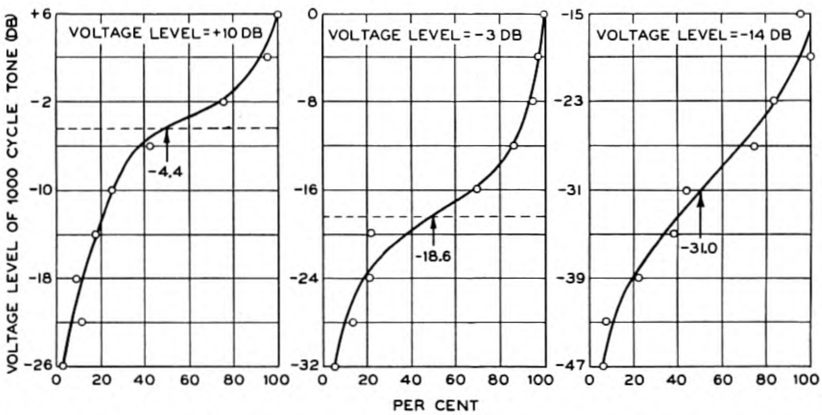


Fig. 17—Percent of observations estimating 1000-cycle tone to be louder than 125-cycle tone.

level at which the 1000-cycle reference, in the opinion of this group of observers, sounded equally loud to the 125-cycle tone.

The testing method adopted was influenced by efforts to minimize fatigue effects, both mental and physical. Mental fatigue and probable changes in the attitude of an observer during the progress of a long series of tests were detected by keeping a record of the spread of each observer's results. As long as the spread was normal it was assumed that the fatigue, if present, was small. The tests were conducted on a time schedule which limited the observers to five minutes of continuous testing, during which time approximately fifteen observations were made. The maximum number of observations permitted in one day was 150.

To avoid fatiguing the ear the sounds to which the observers listened were of short duration and in the sequence illustrated on Fig. 18. The

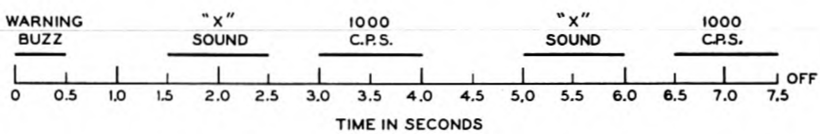
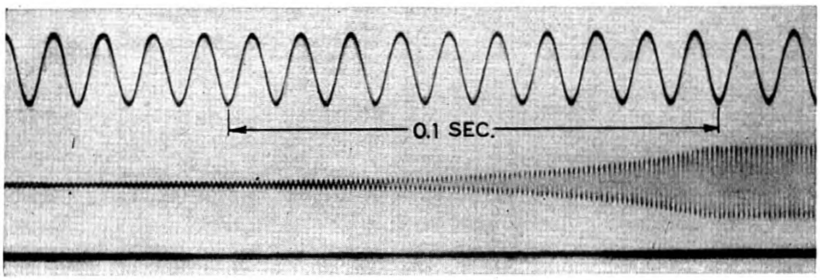


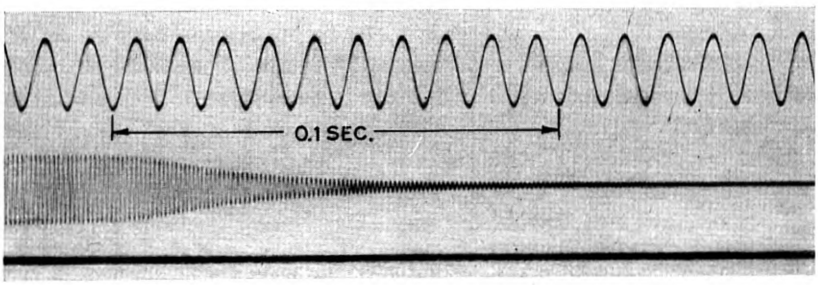
Fig. 18—Time sequence for loudness comparisons.

duration time of each sound had to be long enough to attain full loudness and yet not sufficiently long to fatigue the ear. The reference tone followed the x sound at a time interval short enough to permit a ready comparison, and yet not be subject to fatigue by prolonging the stimulation without an adequate rest period. At high levels it was found that a tone requires nearly 0.3 second to reach full loudness and if sustained for longer periods than one second, there is danger of fatiguing the ear.[13]

To avoid the objectionable transients which occur when sounds are interrupted suddenly at high levels, the controlling circuit was designed to start and stop the sounds gradually. Relays operating in the feedback circuits of the vacuum tube oscillators and in the grid circuits of amplifiers performed this operation. The period of growth and decay was approximately 0.1 second as shown on the typical oscillogram in Fig. 19. With these devices the transient effects were



Growth



Decay

Fig. 19—Growth and decay of 1000-cycle reference tone.

reduced and yet the sounds seemed to start and stop instantaneously unless attention was called to the effect. A motor-driven commutator operated the relays which started and stopped the sounds in proper sequence, and switched the receivers from the reference tone circuit to the sound under test.

[13] G. v. Bekesy, "Theory of Hearing," *Phys. Zeits.* **30,** 115 (1929).

The customary routine measurements to insure the proper voltage levels impressed upon the receivers were made with the measuring circuit shown schematically in Fig. 15. During the progress of the tests voltage measurements were made frequently and later correlated with measurements of the corresponding field sound pressures.

Threshold measurements were made before and after the loudness tests. They were taken on the same circuit used for the loudness tests (Fig. 15) by turning off the 1000-cycle oscillator and slowly attenuating the other tone below threshold and then raising the level until it again became audible. The observers signalled when they could no longer hear the tone and then again when it was just audible. The average of these two conditions was taken as the threshold.

An analysis of the harmonics generated by the receivers and other apparatus was made to be sure of the purity of the tones reaching the ear. The receivers were of the electrodynamic type and were found to produce overtones of the order of 50 decibels below the fundamental. At the very high levels, distortion from the filters was greater than from the receivers, but in all cases the loudness level of any overtone was 20 decibels or more below that of the fundamental. Experience with complex tones has shown that under these conditions the contribution of the overtones to the total loudness is insignificant.

The method of measuring loudness level which is described here has been used on a large variety of sounds and found to give satisfactory results.

Appendix B. Comparison of Data on the Loudness Levels of Pure Tones

A comparison of the present loudness data with that reported previously by B. A. Kingsbury [4] would be desirable and in the event of agreement, would lend support to the general application of the results as representative of the average ear. It will be remembered that the observers listened to the tones with both ears in the tests reported here, while a single receiver was used by Kingsbury.

Also, it is important to remember that the level of the tones used in the experiments was expressed as the number of db above the average threshold current obtained with a single receiver. For both of these reasons a direct comparison of the results cannot be made. However, in the course of our work two sets of experiments were made which give results that make it possible to reduce Kingsbury's data so that it may be compared directly with that reported in this paper.

In the first set of experiments it was found that if a typical observer listened with both ears and estimated that two tones, the reference tone and a tone of different frequency, appeared equally loud, then, making a similar comparison using one ear (the voltages on the receiver remaining unchanged) he would still estimate that the two tones were equally loud. The results upon which this conclusion is based are shown in Table XVI. In the first row are shown the frequencies of the tones tested. Under these frequencies are shown the differences in db of the voltage levels on the receivers obtained when listening by the two methods, the voltage level of the reference tone being constant at 32 db down from 1 volt. Under the caption "Other Reference Tone Levels" similar figures for frequencies of 62 c.p.s. and 2000 c.p.s. and for the levels of the reference tone indicated are given. It will be seen that these differences are well within the observational error. Consequently, the conclusion mentioned above seems to be justified. This is an important conclusion and although the data are confined to tests made with receivers on the ear it would be expected that a similar relation would hold when the sounds are coming directly to the ears from a free wave.

TABLE XVI

COMPARISON OF ONE AND TWO-EAR LOUDNESS BALANCES

A. Reference tone voltage level = − 32 db

Frequency, c.p.s.	62	125	250	500	2000	4000	6000	8000	10,000
Voltage level difference *	−0.5	0	+1.0	−1.0	−0.5	−0.5	+0.5	−3.0	−3.0

B. Other reference tone levels

62 c.p.s.		2000 c.p.s.	
Ref. Tone Volt. Level	Volt. Level Difference *	Ref. Tone Volt. Level	Volt. Level Difference *
−20	+0.5	− 3	0.0
−34	+0.2	−22	+0.3
−57	+2.0	−41	−0.8
−68	−0.5	−60	−0.8
		−79	−6.2

* Differences are in db, positive values indicating a higher voltage for the one ear balance.

This result is in agreement with the point of view adopted in developing the formula for calculating loudness. When listening with one instead of two ears, the loudness of the reference tone and also that of the tone being compared are reduced to oné half. Consequently, if they were equally loud when listening with two ears they must be equally loud when listening with one ear. The second set of data is concerned with differences in the threshold when listening with one ear *versus* listening with two ears.

It is well known that for any individual the two ears have different acuity. Consequently, when listening with both ears the threshold is determined principally by the better ear. The curve in Fig. 20 shows the difference in the threshold level between the average of the better of an observer's ears and the average of all the ears. The circles represent data taken on the observers used in our loudness tests while the crosses represent data taken from an analysis of 80 audiograms of persons with normal hearing. If the difference in acuity when listening with one ear *vs.* listening with two ears is determined entirely by the better ear, then the curve shown gives this difference. However, some experimental tests which we made on one ear acuity *vs.* two ear acuity showed the latter to be slightly greater than for the better ear alone, but the small magnitudes involved and the difficulty of avoiding

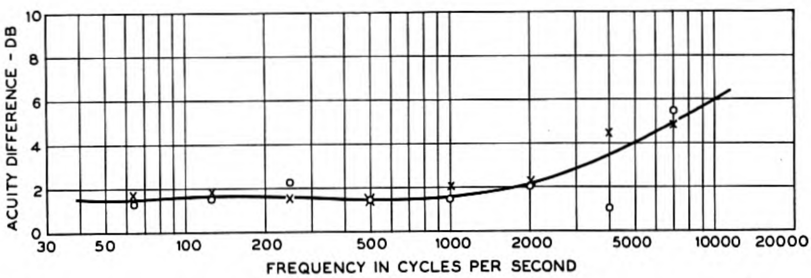


Fig. 20—Difference in acuity between the best ear and the average of both ears.

psychological effects caused a probable error of the same order of magnitude as the quality being measured. At the higher frequencies where large differences are usually present the acuity is determined entirely by the better ear.

From values of the loudness function G, one can readily calculate what the difference in acuity when using one *vs.* two ears should be. Such a calculation indicates that when the two ears have the same acuity, then when listening with both ears the threshold values are about 2 db lower than when listening with one ear. This small difference would account for the difficulty in trying to measure it.

We are now in a position to compare the data of Kingsbury with those shown in Table I. The data in Table I can be converted into decibels above threshold by subtracting the average threshold value in each column from any other number in the same column.

If now we add to the values for the level above threshold given by

Kingsbury an amount corresponding to the differences shown by the curve of Fig. 20, then the resulting values should be directly comparable to our data on the basis of decibels above threshold. Comparisons of his data on this basis with those reported in this paper are shown in Fig. 21. The solid contour lines are drawn through points taken from Table I and the dotted contour lines taken from Kingsbury's data. It will be seen that the two sets of data are in good agreement between 100 and 2000 cycles but diverge somewhat above and below these points. The discrepancies are slightly greater than would be expected from experimental errors, but might be explained

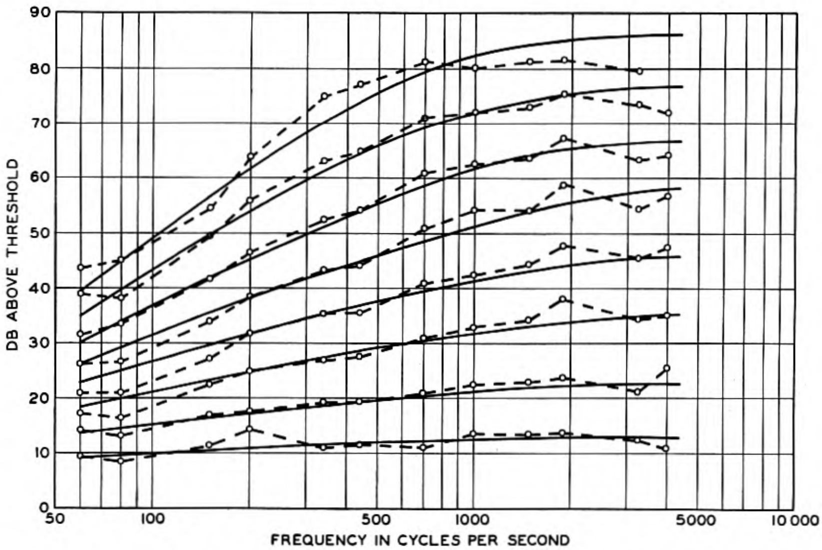


Fig. 21—Loudness levels of pure tones—A comparison with Kingsbury's data.

by the presence of a slight amount of noise during threshold determinations.

Appendix C. Optical Tone Generator of Complex Wave Forms

For the loudness tests in which the reference tone was compared with a complex tone having components of specified loudness levels and frequencies, the tones were listened to by means of head receivers as before; the circuit shown in Fig. 15 remaining the same excepting for the vacuum tube oscillator marked "x Frequency." This was replaced by a complex tone generator devised by E. C. Wente of the Bell Telephone Laboratories. The generator is shown schematically in Fig. 22.

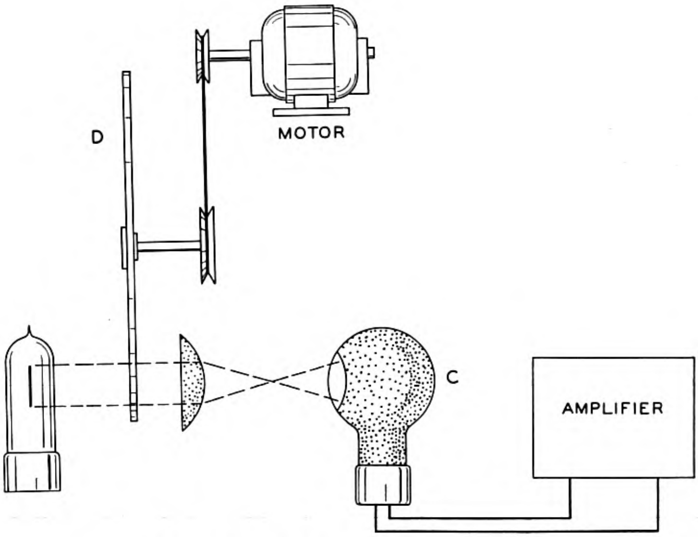


Fig. 22—Schematic of optical tone generator.

The desired wave form was accurately drawn on a large scale and then transferred photographically to the glass disk designated as D in the diagram. The disk, driven by a motor, rotated between the lamp L and a photoelectric cell C, producing a fluctuating light source which

Fig. 23—Ten disk optical tone generator.

was directed by a suitable optical system upon the plate of the cell. The voltage generated was amplified and attenuated as in the case of the pure tones.

The relative magnitudes of the components were of course fixed by the form of the wave inscribed upon the disk, but this was modified when desired, by the insertion of elements in the electrical circuit which gave the desired characteristic. Greater flexibility in the control of the amplitude of the components was obtained by inscribing each component on a separate disk with a complete optical system and cell for each. Frequency and phase relations were maintained by mounting all of the disks on a single shaft. Such a generator having ten disks is shown in Fig. 23.

An analysis of the voltage output of the optical tone generators showed an average error for the amplitude of the components of about $\pm$ 0.5 db, which was probably the limit of accuracy of the measuring instrument. Undesired harmonics due to the disk being off center or inaccuracies in the wave form were removed by filters in the electrical circuit.

All of the tests on complex tones described in this paper were made with the optical tone generator excepting the audiometer, and two tone tests. For the latter tests, two vacuum tube oscillators were used as a source.

Effect of Atmospheric Humidity and Temperature on the Relation between Moisture Content and Electrical Conductivity of Cotton *

By ALBERT C. WALKER

THE data given in this paper show the effect of successive equilibrium humidity cycles on the relation between (*a*) relative humidity and moisture content; (*b*) insulation resistance and relative humidity; and (*c*) insulation resistance and moisture content, for raw and water-boiled cotton at constant temperature (25° C.). These data have been of considerable assistance in explaining the behavior of cotton, particularly the fact that its d.-c. insulation resistance, when measured at some definite test condition,[1] is dependent, to a surprising extent, upon previous treatment, e.g. the manner in which wet cotton is dried, temperature of drying, and the atmospheric conditions to which it is exposed after drying, before being measured under the comparable test condition.

The information secured as a result of this investigation has been valuable in improving the practical methods of inspection used to control the quality of textiles for electrical insulation in telephone apparatus.

Previously it was shown [2] that the relation between the insulation resistance (I.R.) and percentage moisture content (per cent M.C.) of cotton can be expressed by the equation

$$\log \text{I.R.} = -A \log \text{per cent M.C.} + B.$$

It is now known that a single value of the slope A of this linear function does not suffice for all cottons, nor even for one sample of cotton. The slope may have values between 10 and 12 for the same sample depending upon the previous treatment of the cotton. Further, this equation holds only between about 3 per cent and 10 per cent

* This is one of three papers by Walker and Quell, published in the March and April 1933 issues of *The Journal of the Textile Institute.* Abstracts of the other two papers appear in the Abstracts section of this issue of the *Bell System Technical Journal.* In the April 1929 *Bell System Technical Journal* there are two papers by R. R. Williams and E. J. Murphy, and E. B. Wood and H. H. Glenn, respectively, dealing with the problem of textile insulation.

[1] It is the practice to compare the electrical insulating quality of different cotton samples by measuring the d.-c. insulation resistance after bringing the samples to equilibrium with 75 per cent relative humidity at 25° C., or at 85 per cent relative humidity at 37.8° C. (100° F.), equilibrium being approached from a lower humidity.

[2] Murphy and Walker, *J. Phys. Chem.*, **32**, 1761, 1928.

moisture content—corresponding to a range of relative humidity (hereinafter written R.H.) from 15 per cent to 85 per cent at 25° C. Nearly the whole range of moisture adsorption [3] of cotton between dryness and saturation may be characterized by three equations, as follows:

Below 3 per cent moisture content [4]

$$\log \text{I.R.} = -A \text{ per cent M.C.} + B \tag{I}$$

Between 3 per cent and 10 per cent moisture content

$$\log \text{I.R.} = -A \log \text{per cent M.C.} + B \tag{II}$$

Between 10 per cent moisture content and saturation (about 25 per cent M.C.)

$$\log \text{I.R.} = -A \text{ per cent R.H.} + B \tag{III}$$

Different values of A satisfy these equations, depending, as noted above, upon the previous treatment of the cotton and upon the direction of approach to equilibrium; whether this approach is from the dry state (along an absorption cycle), or from the wet state (along a desorption cycle). The experimental data include results of tests on one sample of raw cotton and two of water-boiled cotton. The following tabulation gives some idea of the limiting values of A and the conditions under which they will satisfy the equations:

	Equation I		Equation II		Equation III	
	Raw	Water-boiled	Raw	Water-boiled	Raw	Water-boiled
Absorption....	1.16	No values [5]	10.5 –11	12	0.143	0.111
Desorption....	1.06		9.88–10.15	10.2	0.076	0.075

Experimental Method

Samples of cotton were brought to equilibrium with a flowing stream of air at 25° C., in which the partial pressure of water vapor could be adjusted to any desired value and maintained constant within 0.0115

[3] The word "absorption" is used to denote the taking up of a vapor, "desorption" the giving up of a vapor, and "adsorption" the general process without special indication of gain or loss. The use of these terms implies no assumptions with regard to the mechanism of the processes they denote.

[4] Below 2 per cent M.C., the I.R. of even raw cotton is difficult to measure, since it is above the limiting sensitivity (10^{13} ohms/mm. at 100 volts) of the insulation-resistance bridge used. Further tests are being made on this low range, using a more suitable type of cotton sample.

[5] Difficult to measure water-boiled cotton in the range where Equation I might apply.

mm. This is equivalent to variations of less than 15 parts per million in the water-vapor content of the air, or 0.05 per cent R.H. at 25° C. Insulation-resistance and moisture-content measurements were made at equilibrium [6] for a series of relative humidities in both absorption and desorption cycles on separate samples of the same cotton.

The *moisture content* was determined by mounting about 0.08 gram of cotton, wound in the form of a small skein, on a calibrated quartz-fiber balance, as described by McBain and Bakr.[7] The sensitivity of this spring was 0.03 gram = 1 inch deflection. The deflection caused by moisture adsorption was measured with a cathetometer, calibrated to 0.0001 inch. Measurements were reproducible to 0.0005 inch; thus the moisture adsorbed could be determined to 0.02 per cent.

The *insulation resistance* was measured by mounting 90 threads of cotton, each $\frac{1}{2}$ inch long between metal electrodes, described in a previous communication.[8] This sample weighed about 0.05 gram.

The quartz spring was suspended in a long glass tube mounted within an air thermostat. A metal box with a hard-rubber top on which were mounted the electrodes was also contained in this thermostat. The flowing air streams from the same humidity apparatus were passed through the glass tube and the box in parallel.

A continuous record was obtained of the humidity of the flowing air mixture during each experiment, using an exceedingly sensitive humidity recorder, accurate to 0.05 per cent R.H. at 25° C., and sensitive to changes of but 0.02 per cent R.H. The humidity apparatus and the recorder are both described elsewhere.[9]

Since the humidity apparatus supplied air of fixed absolute humidity, it was essential that constant temperature be maintained in the air thermostat; also that the electrode test box and quartz-spring tube be kept at the same temperature, to insure equilibrium of the samples at the same relative humidity. The *air thermostat* had walls $5\frac{1}{2}$ in. thick, including 3 in. of cork insulation. Copper-constantan thermocouples were mounted in each end of the electrode test-box and in the tube in close proximity to the samples. Efficient circulation of the air within the thermostat, by means of a fan driven from a motor mounted outside the thermostat, together with a sensitive mercury thermo-regulator operating a vacuum tube relay heat control, made it

[6] Below 90 per cent R.H., equilibrium could be practically reached in but two to three hours, using this flowing stream or so-called "dynamic" method. Above 90 per cent, the time for equilibrium increases appreciably, being greater the nearer the test humidity is to saturation. Reference to the data in Table I will show the small differences between two to three hours' exposure and overnight values after 20 hours' exposure.

[7] McBain and Bakr, *Jour. Amer. Chem. Soc.*, **48**, 690, 1926.

[8] April 1929 *B.S.T.J.*, H. H. Glenn and E. B. Wood, Vol. VIII, p. 254.

[9] Walker and Ernst, *Jour. Ind. and Engg. Chem. Analyt. Ed.*, **2**, 134, 1930.

possible to maintain the thermocouples to within 0.01° C. of each other, and the temperature at any point within the thermostat remained constant to at least 0.01° C.

For several years prior to the development of the flowing air stream, or "dynamic" method of testing textiles, insulation-resistance measurements had been made on samples mounted on electrodes in a closed vessel in which 76 per cent R.H. was maintained by saturated NaCl solution. This vessel, in turn, was placed in an air thermostat nearly surrounded by a water bath maintained at 25° C. ± 0.1° C. Since the atmosphere above the salt solution is relatively stationary as compared with that in the flowing stream method, this procedure is defined as a "static" method. A statistical analysis, made by Dr. W. A. Shewhart of these laboratories, on data taken with both the static and dynamic methods, using samples from the same spool of cotton, clearly showed the superiority of the dynamic method.[10]

Experimental Data

Table I contains equilibrium data on moisture content and insulation resistance measurements of raw cotton made at a series of different relative humidities at 25° C., in both absorbing and desorbing cycles. Tables II and III contain similar data for two samples of water-

TABLE I

Moisture Content and Insulation Resistance Data on Raw Cotton in Equilibrium with Constant Atmospheric Humidities during Repeated Absorption and Desorption Cycles at 25° C.

Equilibrium Relative Humidity at 25° C.	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
%	% M.C.	log % M.C.	megohms	log megohms
First Cycle of Increasing Humidity—Absorption				
8.8	2.19	0.340	1.76×10^9	9.25
17.6	3.10	0.491	2.18×10^8	8.34
26.3 (2 hours)	3.76	0.575	2.21×10^7	7.34
26.3 (overnight—20 hours)	3.83	0.584	2.03×10^7	7.31
45.7	5.19	0.72	5.81×10^5	5.76
61.0	6.49	0.813	6.33×10^4	4.80
71.5 (3 hours)	7.61	0.882	8.84×10^3	3.95
71.5 (overnight—21 hours)	7.66	0.885	8.61×10^3	3.94
82.3	9.39	0.973	1.05×10^3	3.02
87.5	11.00	1.041	2.58×10^2	2.41
92.7 (6 hours)	13.95	1.145	41.6	1.62
93.0 (overnight—24 hours)	14.25	1.154	38.0	1.58
99.2	22.30	1.349	5.75	0.76
Saturated air (1 hour exposure)	24.50	1.390	4.17	0.62

[10] This analysis has been published by Dr. Shewhart, as an illustration of testing control in a book, "Economic Control of Quality of Manufactured Product," D. Van Nostrand, 1931. His conclusion regarding this analysis was, "We assume, therefore, upon the basis of this test, that it is not feasible for research to go much further in eliminating causes of variability." Page 21.

First Cycle of Decreasing Humidity—Desorption

93.0	16.90	1.228	15.0	1.18
77.2	10.81	1.034	2.45×10^{2}	2.39
56.0	7.50	0.875	8.90×10^{3}	3.95
36.8	5.32	0.726	3.05×10^{5}	5.48
17.6	3.47	0.540	2.22×10^{7}	7.35
11.1	2.61	0.417	2.25×10^{8}	8.35

Samples dried 20 hours with dry air at 25° C.

Second Cycle of Increasing Humidity—Absorption

26.2	3.90	0.591	1.11×10^{7}	7.05
36.2	4.68	0.670	1.39×10^{6}	6.14
56.5	6.47	0.811	3.87×10^{4}	4.59
71.5 (2 hours)	8.35	0.922	3.34×10^{3}	3.52
72.5 (overnight—18 hours)	8.45	0.927	2.79×10^{3}	3.45
Saturated air (6 hours exposure)	30.00 [11]	1.48	2.64	0.42

TABLE I (*Continued*)

Equilibrium Relative Humidity at 25° C.	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
%	% M.C.	log % M.C.	megohms	log megohms
Second Cycle of Decreasing Humidity—Desorption				
45.0 (2 hours)	6.15	0.79	6.24×10^{4}	4.80
45.0 (overnight—18 hours)	6.08	0.784	6.97×10^{4}	4.84
17.6	3.44	0.537	1.91×10^{7}	7.28
Samples dried 20 hours with dry air at 25° C.				
Third Cycle of Increasing Humidity—Absorption				
26.2	3.88	0.589	9.95×10^{6}	7.00
Desorption from 26.2% to 5% Relative Humidity				
5.0	1.72	0.236	5.67×10^{8}	8.75
Samples removed from apparatus and oven-dried at 80° C. for 20 hours.				
First Cycle of Increasing Humidity—after oven-drying				
45.7	5.07	0.705	4.76×10^{5}	5.68

[11] Under the "saturated" condition in this case, moisture as dew was visible on the cotton.

boiled cotton, designated A and B respectively. These raw and water-boiled samples initially came from the same lot of raw insulating cotton.

The arrangement of the data in these tables shows the sequence in which the equilibrium values were obtained.

On Fig. 1 are plotted curves showing the relations between (*a*) per cent M.C. and per cent R.H., and (*b*) log I.R. and per cent R.H. for the raw cotton data in Table I. Fig. 2 contains a single curve showing the relation between log I.R. and log per cent M.C. for the raw cotton. Fig. 3 contains all three of these different types of curves for the two samples of water-boiled cotton. Since the data for these two water-boiled samples checked with one another so well, only one curve of each type was necessary to express the relations for both samples. Fig. 4 shows the relation between log I.R. and per cent M.C. for only the lower range of the experimental data for raw cotton, since up to about 5 per cent moisture content this relation as expressed by equation I on page 432 appears to hold better than equation II.

Discussion of Experimental Data

Moisture Content-Relative Humidity Data

Exposure of raw cotton to a saturated atmosphere causes a reduction in the area of the moisture content-relative humidity hysteresis loop [12] (Fig. 1). Conversely, no reduction in the area of the loop on successive cycles is observed in the case of water-boiled cotton, perhaps due to this previous water treatment.

Sheppard and Newsome [13] found reductions in the area of this type of hysteresis loop for a treated cotton on successive cycles of exposure to high and low humidities. Our data show—(*a*) no change occurs in the position of the absorption curve for water-boiled cotton during two absorption cycles; (*b*) identical desorption curves for two different water-boiled samples; (*c*) identical desorption curves for raw cotton in three cycles, as well as a suggestion that the third absorption curve (only one point obtained—at 26 per cent R.H.) coincides with the second absorption curve; (*d*) a reduction in area in the raw cotton hysteresis loop on the second absorption cycle; (*e*) this reduced area for the raw cotton differs but little, both in area and location, from the hysteresis loop for the water-boiled cottons.

[12] This type of hysteresis loop in the moisture adsorption properties of cotton has been discussed at length by Urquhart and Williams, *Jour. Text. Inst.*, **15**, T138, 1924; also *Shirley Inst. Mem.*, **3**, 49, 1924.
[13] Sheppard and Newsome, *Jour. Phys. Chem.*, **33**, 1819, 1929.

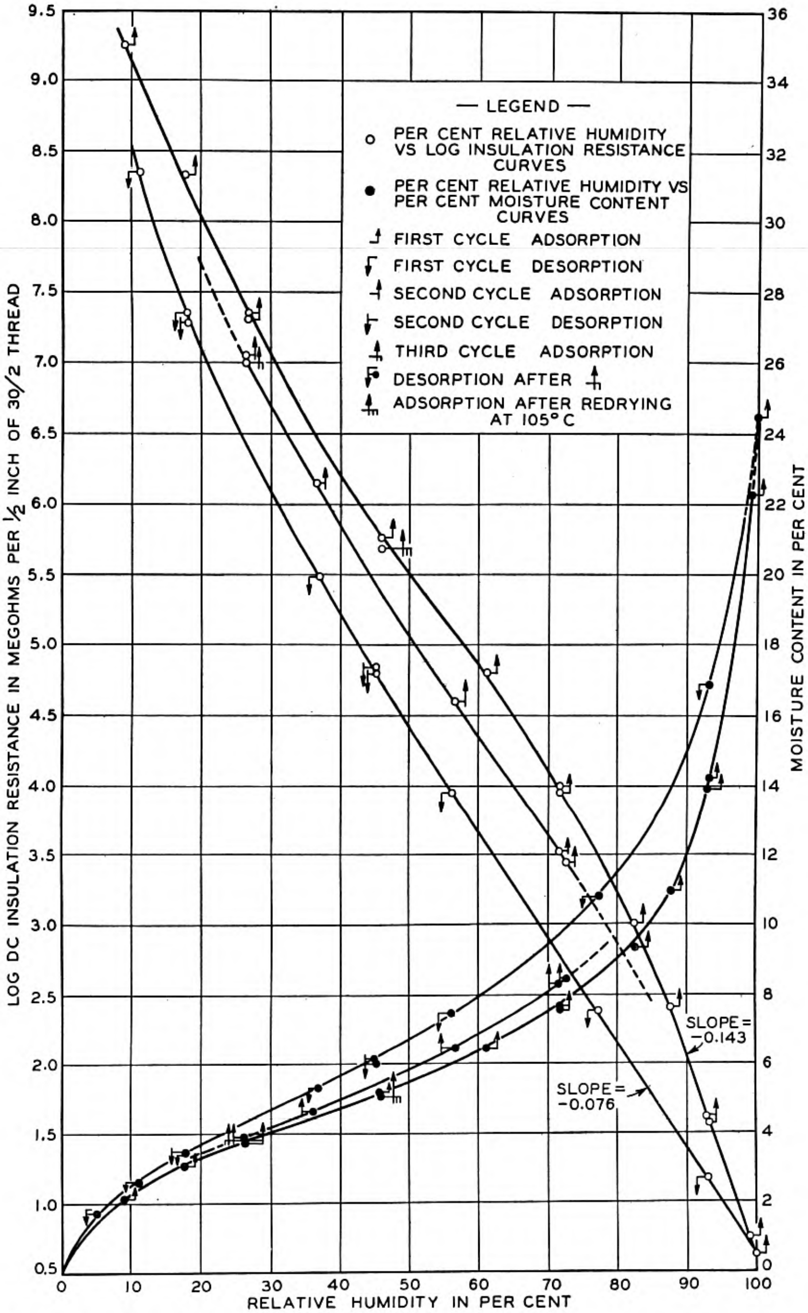


Fig. 1—Relations between relative humidity and the moisture content and log insulation resistance of raw cotton at 25° C.

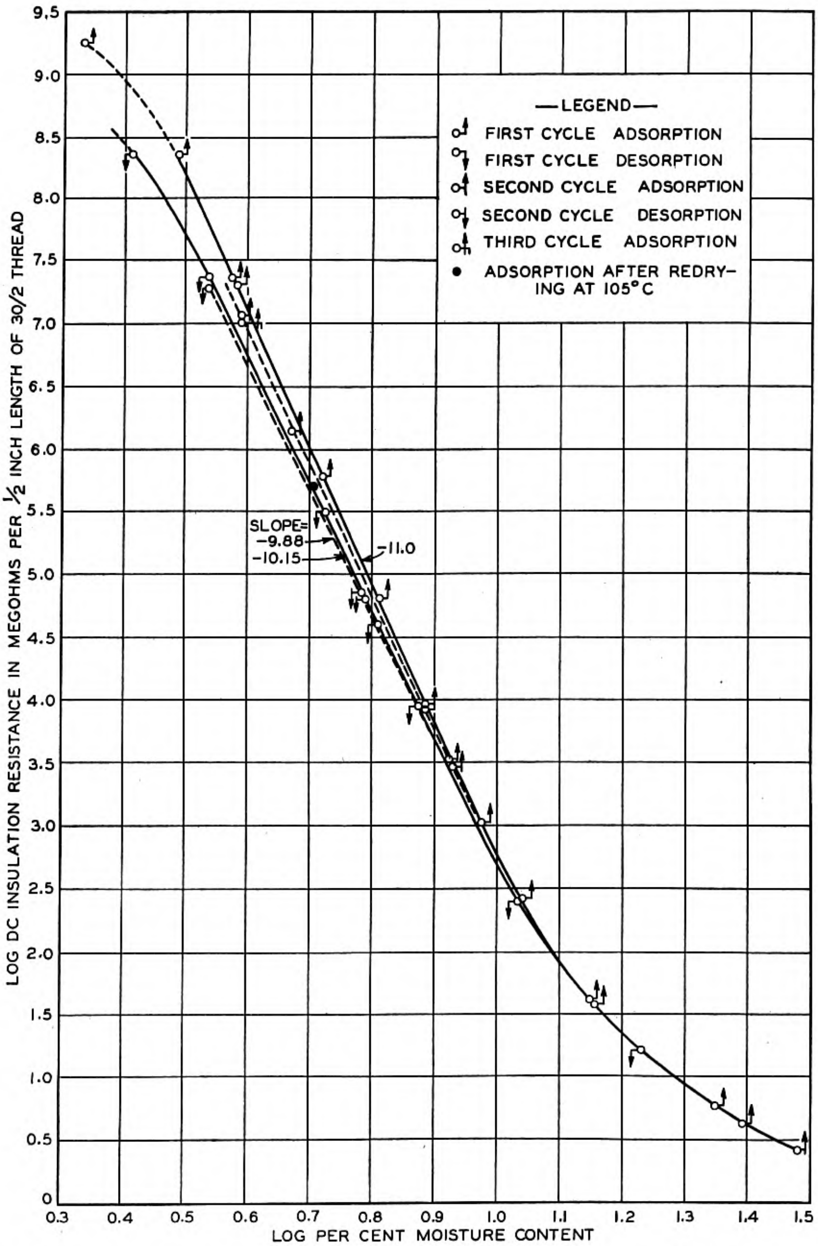


Fig. 2—Relation between log of per cent moisture content and log insulation resistance of raw cotton at 25° C.

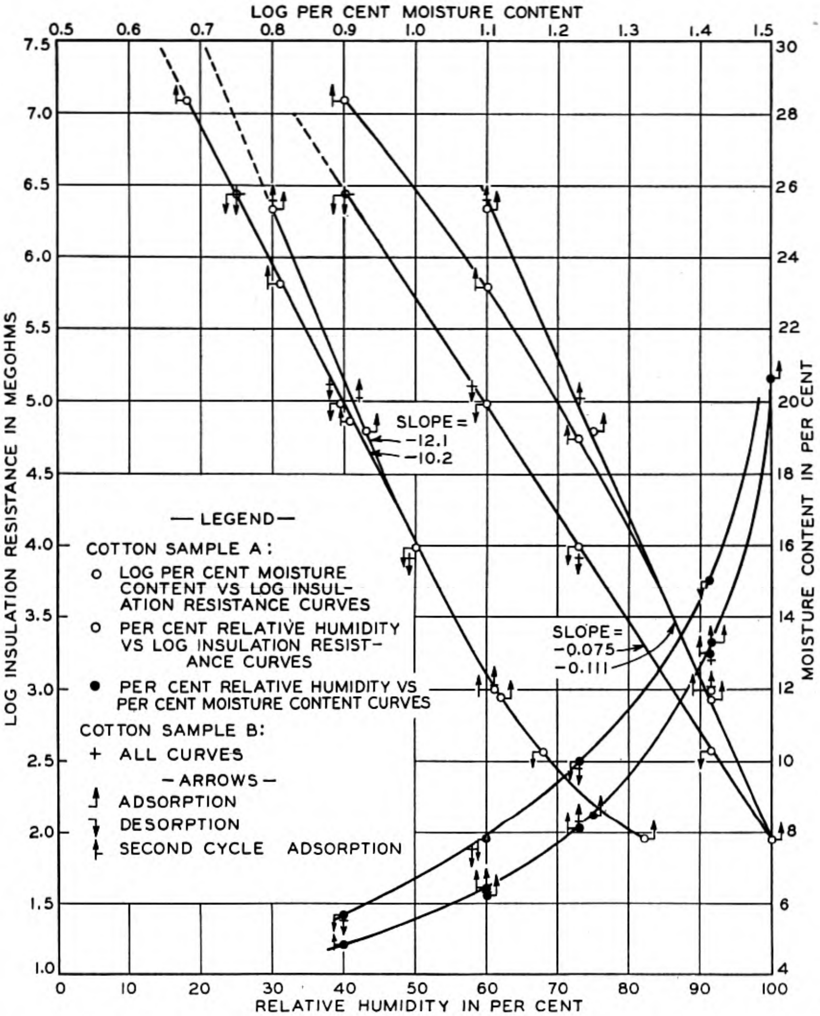


Fig. 3—Relations between relative humidity, moisture content and log insulation resistance of water-boiled cotton at 25° C.

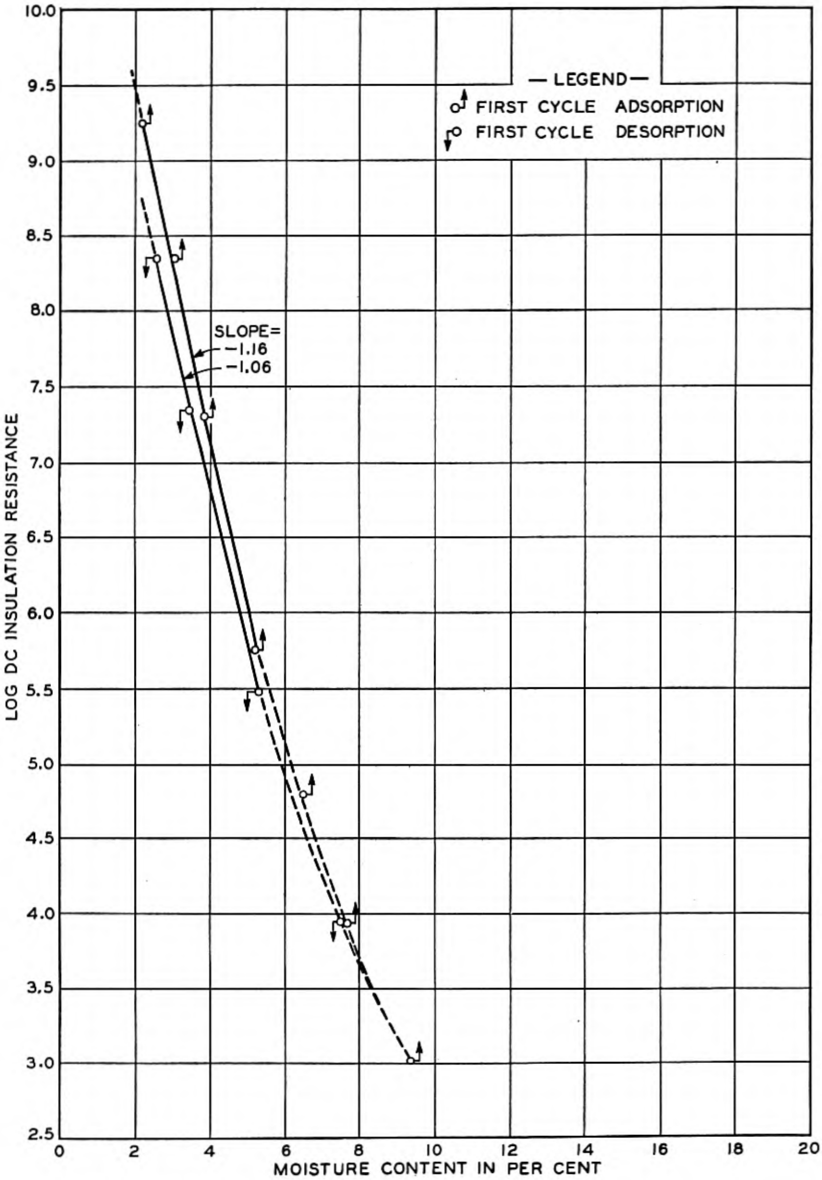


Fig. 4—Relation between per cent moisture content and log insulation resistance of raw cotton at 25° C.

This evidence is considered to indicate the close control of the testing conditions made possible with the dynamic method, and suggests that the decreases in area in the loops obtained by Sheppard and Newsome may be due to small variations in thermostat temperature about a mean value. On absorption this would have the effect of giving too high a moisture content at equilibrium, due to hysteresis; on desorption the equilibrium value would be too low.

TABLE II

MOISTURE CONTENT AND INSULATION RESISTANCE DATA ON WATER-BOILED COTTON IN EQUILIBRIUM WITH CONSTANT ATMOSPHERIC HUMIDITIES DURING ABSORPTION AND DESORPTION CYCLES AT 25° C.

30/2 Cotton—Sample A

Equilibrium Relative Humidity at 25° C.	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
%	% M.C.	log % M.C.	megohms	log megohms
First Cycle of Increasing Humidity—Absorption				
60.0	6.29	0.80	2.21×10^6	6.34
75.0	8.53	0.93	6.3×10^4	4.80
91.5	13.32	1.12	8.93×10^2	2.95
Saturation (20 hours exposure)	20.70	1.32	9.35×10	1.97
First Cycle of Decreasing Humidity—Desorption				
91.5	15.00	1.18	3.80×10^2	2.58
73.0	10.05	1.00	9.75×10^3	3.99
60.0	7.85	0.895	9.46×10^4	4.98
40.0	5.62	0.75	2.77×10^6	6.44
Samples dried 20 hours with dry air at 25° C.				
Second Cycle of Increasing Humidity—Absorption				
40.0	4.80	0.68	1.25×10^7	7.097
60.0	6.45	0.81	6.45×10^5	5.81
73.0	8.16	0.91	5.95×10^4	4.75
91.5	13.03	1.11	1.00×10^3	3.00

Insulation Resistance-Relative Humidity Data

Figs. 1 and 3 show hysteresis loops in the log I.R.—per cent R.H. curves, for both raw and water-boiled cotton. Hysteresis loops in this relation were shown in a previous paper[2] but no evidence was available to show the effect on the loop area of exposure of the

[2] loc. cit.

textile to air saturated with water vapor. From the evidence given in this paper it is seen that exposure to saturated air causes a reduction in the hysteresis loop area for both raw and water-boiled cotton. This behavior is in contrast to the moisture content-relative humidity relation in which a reduction in loop area is observed for raw, but not for water-boiled cotton.

Between 11 per cent moisture content (about 88 per cent relative humidity) and saturation, the log I.R.—per cent R.H. relation appears to be nearly linear for raw cotton, and on the desorption curve the relation is linear down to about 45 per cent R.H. For water-boiled

TABLE III

MOISTURE CONTENT AND INSULATION RESISTANCE DATA ON WATER-BOILED COTTON IN EQUILIBRIUM WITH CONSTANT ATMOSPHERIC HUMIDITIES DURING ABSORPTION AND DESORPTION CYCLES AT 25° C.

30/2 Cotton—Sample B

Equilibrium Relative Humidity at 25° C.	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
%	% M.C.	log % M.C.	megohms	log megohms
First Cycle of Increasing Humidity—Absorption				
73.0	8.33	0.92	1.08×10^5	5.03
60.0	6.33	0.80	2.565×10^6	6.41
91.5	12.87	1.11	1.05×10^3	3.02
Exposed to air at 100% R.H. overnight—no measurements taken.				
First Cycle of Decreasing Humidity—Desorption				
73.0	9.86	0.99	8.33×10^3	3.92
58.0	7.57	0.88	1.31×10^5	5.12
40.0	5.60	0.748	2.78×10^6	6.44

cotton, this relation appears to be substantially linear over the full range investigated, from 60 per cent R.H. to saturation on the absorption curve, and from saturation down to about 40 per cent R.H. on the desorption cycle. Curiously, the second absorption cycles for both raw and water-boiled cotton do not exhibit such a linear relation, although in the range above 90 per cent R.H. it is possible that these second absorption curves join the initial absorption curves and become linear in the upper range.

These curves emphasize the necessity for systematic treatment of textiles in making electrical measurements under definite humidity conditions, since the hysteresis in the per cent R.H.—per cent M.C.

curves indicates that similar hysteresis in the log I.R.—R.H. curves is due to adsorption of different amounts of moisture by cotton, even when exposed to the same relative humidity. The amount of moisture adsorbed is dependent upon the direction from which equilibrium is approached.

Unfortunately, the behavior of cotton is still further complicated, so that additional precautions must be taken in measuring its electrical properties.

The difference in the effect of saturation on the area of the hysteresis loops for raw and water-boiled cotton as shown by the log I.R.—per cent R.H. and per cent R.H.—per cent M.C. curves suggests that some change in structure of cotton occurs when it absorbs much moisture, and this change in structure has a more or less permanent effect on the subsequent behavior of the material. Verification of this suggestion is found in the log I.R.—log per cent M.C. relation which will now be discussed. The study of this log relation has led to many improvements in methods now employed in the fundamental investigation of the electrical properties of cotton and in inspection methods employed in the commercial purification of cotton for electrical purposes.

Insulation Resistance-Moisture Content Data

The curves expressing the relation between log I.R.—log per cent M.C. are shown, in Figs. 2 and 3, to be curved, and not linear over the whole range as suggested in an earlier paper.[2] The data on raw cotton extends over the wider range, and the curve appears to be sigmoid in shape, exhibiting curvature above 10 per cent and below 3 per cent moisture content. Only in the middle range between these moisture content limits is the curve sufficiently linear so that equation II applies. The accuracy of the curve below about 5 per cent M.C. progressively decreases, due to difficulties in measuring the extremely high resistances, and about all that can be said of this range at present is that the log I.R.—per cent M.C. relation expressed by equation I, appears to fit the data better than the log I.R.—log per cent M.C. relation as expressed by equation II.

The definite curvature above 10 per cent M.C., not observed previously,[2] was found through the use of the dynamic method and the measurement of insulation resistance and moisture content values simultaneously on similar samples of cotton taken from the same supply.[14]

[14] In the vicinity of saturation, an effect similar to polarization can cause errors in the measurement of insulation resistance. The errors result in high insulation resistance values, accentuating the curvature of the curve above 10 per cent moisture

In the range where equation II is applicable the relation is seen to be a family of convergent lines with slopes (the constant A in this equation) having values between 10 and 12. These convergent lines focus at about 10 per cent M.C. (log per cent M.C. = 1).[15] The actual value of the slope A in any test depends upon several factors. It is primarily dependent upon the previous treatment of the cotton. Water-boiled cotton which has been dried from the wet state at high temperature in such a manner as to secure a high I.R. for a given moisture content, in consequence, gives a line with maximum slope. Exposure to high humidities, or saturation of the cotton with water vapor causes the subsequent desorption and absorption equilibrium values to lie on a line of less slope. In the case of raw cotton, the more moisture absorbed by the cotton from a saturated atmosphere, the lower is the desorption value of A; its lower limit appears to depend to some extent upon the time of exposure and the amount of moisture absorbed. (Note the difference in the desorption slope after the first and second exposure of the raw cotton to saturated air. After the first cycle with 24.5 per cent maximum moisture content, $A = 10.15$; after the second with 30 per cent M.C., $A = 9.88$.[16] This difference is greater than experimental error.)

Raw cotton shows a distinct difference from water-boiled cotton in one respect. On the second absorption cycle the slope A has a value

content. This effect is not readily detectable, using the slow-period H.S. type Leeds and Northrop galvanometer. When first found, it was assumed that the entire curvature of the curve above 10 per cent M.C. was due to this effect, but such was not the case. The effect is not true polarization, but is simply due to electrical heating. Above 90 per cent relative humidity for raw cotton and above 98 per cent R.H. for washed cotton, the measuring current, using 100 volts potential is sufficient to heat the cotton appreciably. This I^2R loss can raise the textile temperature about 0.1° C. at 90 per cent R.H., and about 10° C. at saturation for raw cotton. These temperature rises were measured, using thermocouples of No. 40 wire braided into the threads of textile mounted on the electrodes. The heating effect causes evaporation of moisture from the cotton, thus raising the insulation resistance.

All measurements in this paper above 75 per cent R.H. for raw cotton and above 90 per cent R.H. for washed cotton were made with a special micro-ammeter having a period of but 0.8 second, as compared with the period of the H.S. type galvanometer of about 40 seconds. The temperature rise at saturation does not become evident for at least three seconds after voltage application. Until this short interval has elapsed the micro-ammeter gives a steady reading identical with the instantaneous value, and as the thermocouple records increasing temperature the meter deflection drops.

[15] This behavior is a hysteresis effect of a somewhat different character from that observed in the two relative humidity relations previously discussed, since in this case the effect is independent of relative humidity, and appears to be related to the distribution of moisture in the cotton and to the manner in which this moisture is held by the cellulose. This will be discussed somewhat more fully later.

[16] The value of 24.5 per cent M.C. does not necessarily indicate a true saturation value, but only a M.C. after exposure to a definite saturated atmosphere for one hour. The 30 per cent value probably represents some value above the critical saturation point at exactly 100 per cent R.H. (which would be exceedingly difficult to obtain), since actual deposits of dew were visible on the sample.

TABLE IV

EFFECT OF HIGH RELATIVE HUMIDITY (88%) AT DIFFERENT TEMPERATURES ON THE INSULATION RESISTANCE OF COTTON AT 75% RELATIVE HUMIDITY AND 25° C.

Sequence of Equilibrium Conditions	Insulation Resistance of Cotton in Kilomegohms per ½-in. Thread														
	Washed Cotton Samples[17]									Raw Cotton Samples[18]					
	1	2	3	4	5	6	7	8	Avg.	1	2	3	4	Avg. (a)	Average (b) Exposed to 88% R.H.
75% R.H.—25° C.	73	80	80	90	100	102	100	159	100	4.6	4.8	4.7	4.7	4.7	—
88% R.H.—22° C.	9.0	9.8	11.3	12.0	12.0	12.5	9.5	15.0	11.0	—	0.48	0.47	—	—	0.48
Dried overnight															
75% R.H.—25° C.	46	50	57	60	60	65	57	94	61	4.5	3.0	2.9	4.6	4.6	2.95
88% R.H.—30.2° C.	2.4	2.1	2.1	2.6	3.0	2.6	2.6	3.8	2.6	—	0.136	0.138	—	—	0.137
Dried overnight															
75% R.H.—25° C.	30	31	34	36	36	41	36	58	36	4.3	1.95	1.95	4.3	4.3	1.95
88% R.H.—38° C.	1.5	0.84	0.78	1.06	0.96	1.53	1.90	2.3	1.11	—	0.09	0.09	—	—	0.09
Dried overnight															
75% R.H.—25° C.	22	23	24	29	26	29	31	42	28	4.6	1.7	1.7	4.6	4.6	1.7
88% R.H.—22° C.	4.3	4.3	4.7	5.8	5.4	6.3	6.0	7.3	5.5	—	0.43	0.36	—	—	0.40
Dried overnight															
75% R.H.—25° C.	34	33	34	34	32	38	38	50	37	4.5	2.1	1.8	4.5	4.5	1.95

[17] These samples were washed at 40° C. in accordance with the procedure described in the paper, "Naturally Occurring Ash Constituents of Cotton," by Walker and Quell.

[18] Two of these raw cotton samples (1 and 4) were used as controls to check the reproducibility of the 75% humidity condition. They were *not* exposed to the 88% humidity conditions. Therefore the averages of 1 and 4 are given under (a). The averages of the other two (2 and 3), which were exposed to the sequence of 88% conditions, are given under (b).

intermediate between the initial absorption and desorption slopes, thus indicating some reversibility in the properties of the cotton which determine these slopes, due to the drying effect after the initial desorption test. Water-boiled cotton does not show this effect, the slope of the second absorption curve being identical with that of the initial desorption curve, *under the conditions of drying used for these tests.* This behavior is consistent with some experiments made to determine if the initially high insulation resistance observed in some cases with water-boiled cotton could be restored by some simple means after the resistance had been adversely affected by exposure to high atmospheric humidities.

In the course of some I.R. tests made on washed cotton the control samples of raw cotton used to check each I.R. experiment to assure the same humidity and temperature conditions were found to have suddenly changed from 4.5 kilomegohms—their normal value under the test conditions—to 1.8 kilomegohms under these conditions. These controls had been exposed to atmospheric humidity conditions of 83 per cent R.H. at 32° C., while a new set of washed cotton samples were being prepared for test. Since it was particularly desirable to continue the use of the same control samples, an attempt was made to restore them to their original conditions by drying. Air at less than 0.1 per cent R.H. at 25° C., was passed over these samples for 40 hours at room temperature. When subsequently measured their resistances had increased from 1.8 to 2.9. Further drying for 48 hours at 105° C. caused a further gain of but 0.1 kilomegohm. Conversely, similar tests on washed cotton showed no improvement. A bundle of washed cotton was dried at 105° C. Instead of giving an I.R. of between 100 and 400 kilomegohms, normal for other similarly washed and dried samples, the resistance was but 23 kilomegohms. Chemical analyses of this cotton gave no indication that this low value was due to electrolytic contamination. Neither redrying of this cotton in a vacuum oven at 80° C., nor drying in an air-oven at 105° C., gave any improvement; in fact the resistance after such redrying was but 18 kilomegohms.

However, this washed cotton was greatly increased in I.R. by simply rewetting with excess water and drying rapidly at 105° C.[19]

From this discussion of the data it is seen that three types of linear equations may be used to express fairly accurately the relation between insulation resistance and the moisture-absorbing properties of cotton over a range of atmospheric relative humidity from saturation down

[19] Samples *A* and *B* used to secure the data in Tables II and III were from this test. After rewetting and oven-drying at 105° C., sample *A* gave 108 kilomegohms and *B* gave 63 kilomegohms at 75 per cent R.H.—25° C.

to nearly dryness. These equations, with the respective ranges of relative humidity (and, therefore, of moisture content) over which each is significant, are given on page 432.

It is concluded that exposure of cotton to high atmospheric humidity causes a change in the gel structure due to absorption of moisture, since the insulation resistance of the material as measured at some comparable condition (75 per cent R.H. at 25° C.) is less after such high humidity exposure than before, even if the cotton is well dried before testing.

The *temperature* of such exposure to high atmospheric humidity also affects the subsequent electrical properties of the cotton. Data to show this temperature effect are given in Table IV.

Temperature Effects

Effect of Temperature at High Humidity on I.R. of Air-dried Cotton

Table IV contains the results of a series of tests on the I.R. of samples of raw and washed cotton which were exposed to several cycles of high humidity and dry air, each cycle being as follows:

(*a*) Equilibrated and measured at 75% R.H.—25° C.
(*b*) Equilibrated and measured at 88% R.H.—at *t*° C.
(*c*) Dried for 16 hours with a stream of dry air at 25° C.

This cycle was repeated four times, the only difference in each case being the temperature (*t*° C.) at which the 88 per cent R.H. equilibrium tests were made. These temperatures were successively—22°, 30.2°, 38°, and 22° C. In all, eight samples of washed cotton and four samples of raw cotton were used in the test. Two of the raw cotton samples (1 and 4) were not exposed to the 88 per cent humidity conditions, but were used as control samples to check the reproducibility of the 75 per cent humidity conditions in each cycle.[20]

Table V is a condensation of Table IV. The decreases in insulation

[20] Five measurements each were made on these two control samples during the course of the test, giving a mean value of 4.52 kilomegohms, with a standard deviation of but 0.13 kilomegohms.

The differences in the initial values of I.R. for the eight washed samples are not due to lack of control, either in the method of washing or in the method of testing, but to actual differences in the equilibrium moisture contents. For example—sample 1 gave 73 kilomegohms initially, and sample 6 gave 102 kilomegohms. Their respective moisture contents, under the test conditions, were 8.17% and 8.00%.

Using Equation II, and with the constant $A = 10$, the values of B were calculated in this equation as 13.99 and 14.05 respectively for samples 1 and 6. Assuming these samples to be of equal purity, since they were washed in an efficient manner,[21] it is reasonable to take $B = 14.03$ for both samples. From this value of B, the I.R. of sample 1 was calculated at a moisture content of 8.00 per cent, giving 98 kilomegohms, a satisfactory check with sample 6 at the same moisture content.

[21] Walker & Quell, *Jour. Text. Inst.* **24**, T141, 1933.

resistance of both raw and washed cotton when measured at 75 per cent relative humidity and 25° C., *after* exposure to the 88 per cent relative humidity conditions and dried,[22] are given in percentage of the *initial* 75 per cent—25° C. insulation resistances.

TABLE V

PERCENTAGE REDUCTION IN THE INSULATION RESISTANCE OF RAW AND WASHED COTTONS AT 75 PER CENT RELATIVE HUMIDITY AND 25° C., *after* SUCCESSIVE EXPOSURES TO 88 PER CENT RELATIVE HUMIDITY AT t° C.

Temperature (t° C.) of the Successive 88% R.H. Cycles	% Reduction in Insulation Resistance at 75%—25° C. after each 88% R.H. Cycle	
	Washed	Raw
22° C.	39%	37%
30.2° C.	64%	58.5%
38.0° C.	72%	64.5%
22° C.	63%	59.5%

Exposure of cotton to high humidity (in this case 88 per cent) alters the properties of the material in such a way that its insulation resistance when subsequently measured at 75 per cent relative humidity and 25° C., is lower than the insulation resistance measured at the 75 per cent condition before such exposure to 88 per cent humidity. This decrease in insulation resistance observed at 75 per cent humidity and 25° C., becomes progressively greater the higher the temperature of the 88 per cent humidity exposure, but on again exposing the cotton to 88 per cent humidity at the reduced temperature of 22° C., after exposure at 38° C., the insulation resistance subsequently measured at 75 per cent humidity and 25° C., is greater than after the 88 per cent—38° exposure, but less than when measured at this condition after the original exposure to 88 per cent humidity and 22° C., thus indicating that some reversal occurs in the temperature effect.

The fact that in each test the percentage reduction is of the same order of magnitude for raw and washed cotton, suggests that the effect is structural and not related to the quantity of electrolytic impurities which may be present.

An important feature of the data recorded in Table IV is that the insulation resistance of washed cotton is reduced by exposure to 88 per cent R.H. A natural question is—What would be the resistance of this cotton if exposed to 100 per cent R.H. instead of 88 per cent, or brought directly to equilibrium with 75 per cent R.H. at 25° C., from the wet state without oven-drying? Tests have been made to determine these points. Washed cotton, dried at 105° C., then con-

[22] The samples were dried with a stream of very dry air at 25° C. after each exposure to the 88 per cent humidity conditions to avoid the hysteresis effect, which would occur if the samples were brought back to the 75 per cent humidity condition directly from the higher humidity. Before starting the test all samples were similarly dried.

ditioned at 100 per cent R.H., gave an I.R. when tested at 75 per cent R.H. at 25° C., of 25 kilomegohms.[23] Its insulation resistance on being brought directly from the wet state to 75 per cent R.H. at 25° C., was but 3.7 kilomegohms, being in this case lower than the resistance of raw, unwashed cotton in Table IV. Of course, if the raw cotton could be wet with water without undergoing any change due to reduction in ash content, no doubt its resistance would be much lower than that of similarly treated water-washed cotton, since this effect appears to be structural and certainly is not dependent upon electrolytic impurities.

Effect of Temperature of Drying Wet Cotton on its Insulation Resistance

The higher the temperature at which wet, water-boiled cotton is dried, the higher is its insulation resistance. Such cotton, dried at 105° C., 120° C., and 162° C., from the wet state, gave 139, 171, and 201 kilomegohms respectively, when subsequently equilibrated at 75 per cent R.H. at 25° C.

THEORY

The most important fact to be derived from these experimental data is that cotton may have a range of insulation resistance values for any single moisture content over at least the average atmospheric humidity range, from about 15 to 85 per cent R.H. Another interesting fact is that the insulation resistance of cotton when measured at definite test conditions depends to a surprising extent upon the previous exposure of the material to prevailing atmospheric humidity and temperature conditions, prior to such tests.

This behavior suggests that the absorption of appreciable quantities of moisture causes changes in the cotton structure, which affect the mechanism of current conduction. This change in structure, no doubt a result of swelling, an effect investigated by Collins,[24] appears to be a difficultly reversible alteration in the colloidal gel structure of the cellulose, even after subsequent removal of the moisture by drying. These effects, rather small to be detected by ordinary methods, are revealed by the extremely sensitive electrical tests, since very small changes in moisture content cause large changes in insulation resistance.

Since the substitution of acetyl for hydroxyl groups in cellulose is accompanied by a marked reduction in the moisture adsorption,[25]

[23] This oven-dried material gave 80 kilomegohms when not exposed to the 100 per cent R.H. before test.

[24] Collins, *Jour. Text. Inst.*, **21**, T311, 1930.

[25] Wilson and Fuwa, *Jour. Ind. and Engg. Chem.*, **14**, 913, 1922. (This lower moisture adsorption of cellulose acetate has been observed in our own experiments. See also reference [13].)

it appears likely that adsorption of moisture is largely a function of free hydroxyl groups. From our data it appears that when wet cotton is dried rapidly at high temperatures, the internal or micelle surface contains a minimum of hydroxyl groups. As the cotton is permitted to absorb more and more moisture, the hydroxyl groups which were oriented into the interior of the micelles by the drying process where their hygroscopic property is, in effect, neutralized by attraction of associated molecules, are attracted to the surface to hold the absorbed moisture. On drying, these hydroxyl groups do not return readily to the interior and a greater number of water molecules are held at any relative humidity, thus accounting for the normal hysteresis effect observed in the moisture content-relative humidity relation.

Practically all of the experimental data discussed in this paper were secured during 1928 and 1929, and the above theory was proposed at that time. Apparently at about the same time Urquhart questioned the explanation offered some years previously by Urquhart and Williams [26] to account for hysteresis in the moisture relations of cotton, depending upon a modification of the Zsigmondy pore theory. In June 1929 [27] Urquhart proposed a theory comprising the essential features of the orientation of hydroxyl groups as offering a better explanation than the pore theory for the moisture-adsorbing properties of cotton. The general outline just given in connection with the study of the electrical properties of cotton is much the same as the more complete theory discussed by Urquhart.

However, further consideration of our experimental data led to the conclusion that neither the pore theory nor the orientation of hydroxyl groups completely accounts for the hysteresis effect in the log I.R.—log per cent M.C. relation.

As mentioned above, rapid drying of wet cotton under proper conditions is assumed to give internal surfaces containing a minimum of hydroxyl groups. This idea can be qualified as follows: Either such drying conditions are conducive to the presence of a minimum of hydroxyl groups on the internal surfaces, or they are conducive to a *less uniform distribution* of these groups on these internal surfaces.

Consequently, on initially absorbing moisture from such a dried condition, the moisture associated with hydroxyls will not be uniformly distributed and the conduction of current through the cotton along these internal surfaces will be somewhat *discontinuous*. On desorption from saturation, moisture will be removed in a more regular

[26] Urquhart and Williams, *Jour. Text. Inst.*, **15**, T433, 1924; also *Shirley Inst. Mem.*, **3**, 197, 1924.
[27] Urquhart, *Jour. Text. Inst.*, **20**, T125, 1929.

fashion from more uniformly distributed hydroxyls, and therefore on any descending curve the conduction of current can be considered to be along more continuous paths. This difference in continuity of moisture paths is sufficient to account for high insulation resistance values on absorption and low values on desorption curves, for each equilibrium moisture content. The actual insulation resistance in any given case depends upon the degree of continuity of such moisture paths and this in turn depends upon the previous treatment of the material.

Also it seems reasonable to consider that some of the properties of cotton under discussion may be explained to better advantage by the pore theory initially proposed by Urquhart and Williams,[26] since it does not appear that all of the moisture which saturated cotton can absorb is necessarily associated with hydroxyl groups. In considering the pore theory, high insulation resistance values during absorption can be accounted for by a blocking effect of the pore entrances by a few water molecules. This pore blocking effect suggested by Peirce[28] would cause greater discontinuities in moisture paths through the cotton, and therefore higher insulation resistance for a given moisture content.

Since it is planned to discuss this theory more in detail in a separate paper when experimental data now being secured are available, only the above brief outline is given at this time.

Acknowledgments are made to Mr. M. H. Quell, Mr. H. S. Davidson, and Mr. G. E. Kinsley for their valuable assistance in securing the data reported in this paper.

[28] Peirce, *Jour. Text. Inst.*, **20**, T133, 1929.

Classification of Bridge Methods of Measuring Impedances *

By JOHN G. FERGUSON

An analysis is made of the requirements for satisfactory operation of the simple four-arm bridge when used for impedance measurements. The various forms of bridge are classified into two major types called the ratio-arm type and the product-arm type, based on the location of the fixed impedance arms in the bridge. These two types are subdivided further, based on the phase relation which exists between the fixed arm impedances. Eight practical forms of bridges are given, three of them being duplicate forms from the standpoint of the method of measuring impedance. These bridges together allow the measurement of any type of impedance in terms of practically any type of adjustable standard. The use of partial substitution methods and of resonance methods with these bridges is discussed and several methods of operation are described which show their flexibility in the measurement of impedance.

Introduction

BRIDGE methods have been used for the measurement of impedance from the very beginning of alternating current use. In fact, the history of the impedance bridge dates back to the earlier bridges developed for the measurement of direct current resistance. While some objection may be raised to this method of measurement on the count that it is not direct indicating, in the sense that an ammeter or voltmeter is, this has been more than offset by the high accuracy of which it is capable. Bridge methods of measuring impedance have accordingly continued to hold a high place in the field of electrical measurements and except perhaps at the higher radio frequencies are considered supreme for this purpose over the whole frequency range, where high accuracy is the principal requirement.

The peculiar advantages of the bridge method are most evident where emphasis is laid on the circuit characteristics rather than on power requirements. In power engineering it may be more logical to make measurements in terms of current, voltage, and power, since these are the quantities of immediate interest. In communication engineering, however, where design is based for the most part on circuit characteristics, and power considerations are only of secondary interest, it is natural that bridge methods, which furnish a direct comparison of these circuit characteristics should be generally preferred.

* Presented at Summer Convention of A.I.E.E., Chicago, Illinois, June, 1933. Abstracted in June 1933 *Elec. Engg.*

Due to the wide field of usefulness and great flexibility of the impedance bridge, a very large amount of development work has been done and a considerable amount of literature has been published covering various types and modifications. In fact, the subject has become so broad and the information so voluminous that the engineer who has not specialized in the subject has every excuse for a feeling of considerable confusion when he finds it necessary to make a choice among the numerous circuits available. Perhaps the greatest single obstacle to a still more extensive use of the impedance bridge in industry is this very multiplicity of types combined with a rather complete lack of any practical guide for the engineer who is interested principally in the measurement itself and looks on the bridge simply as a means to this end.

Very little information is available as to the relative merits of the various types of bridges, the great majority of published articles being confined to a description of a particular circuit used by the author for a particular purpose.

The present article furnishes a comparison of the relative merits of the large number of circuits which are available for making the same measurement and should serve as a guide to the engineer who is more interested in results than in acquiring a broad education in bridge measurements. An outline is given of the fundamental requirements which must be met by bridges used for impedance measurements, and a classification is made which serves as a help in the choice of a bridge for any particular type of measurement. The relative merits of the simpler types of bridge are discussed from the standpoint of the measurement of both components of an impedance, particularly with reference to measurements in the communication range of frequencies from about 100 to 1,000,000 cycles. Where only the major component of an impedance is desired, for instance where only the inductance of a coil or the capacitance of a condenser is desired, the requirements are not so severe and many forms of bridges may be used which are not suitable for the purpose here outlined. Bridges are also used to a large extent for other purposes than impedance measurements, such as for frequency measurements. These applications will not be considered here.

The General Bridge Network

Any bridge may be considered as a network consisting of a number of impedances which may be so adjusted that when a potential difference is applied at two junction points, the potential across two other junction points will be zero. For this condition, there are relations between certain of the impedances which enable us to evaluate one of them in terms of the others. Thus the bridge is essentially a method of comparing impedances. The impedances of the bridge may consist of resistance, capacitance, self and mutual inductance, in any combinations, and they may actually form a much more complicated network than the simple circuit shown in Fig. 1. Consequently, the number of

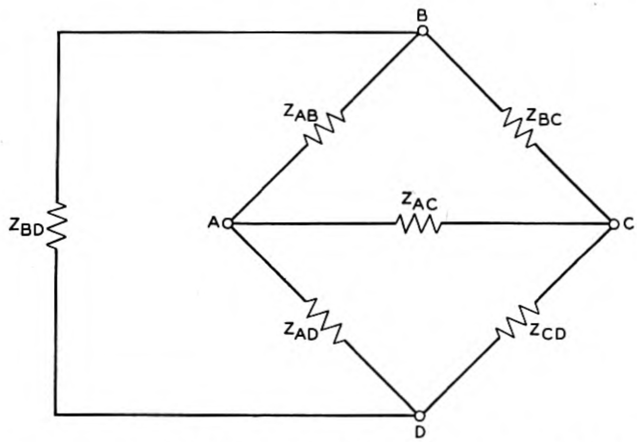


Fig. 1—Schematic of the impedance bridge reduced to its simplest form.

different bridges which can be devised for the measurement of impedances is extremely large. However, since only four junction points are significant, any bridge circuit may be reduced to a network of six impedances connected between these four points, as shown in Fig. 1. These impedances are direct impedances, that is, there are no mutual impedances between them.

If a potential is applied at BD and the balance condition is that the potential be zero across AC, then the points BD are called the input or power source terminals and the points AC are called the output or detector terminals. The impedances Z_{BD} and Z_{AC} then act simply as shunts across the power source and detector respectively and do not affect the balance relation. The balance is not affected if the power source and detector are interchanged in a bridge reduced to this simple form and hereafter no distinction will be made in this respect.

After the bridge has been reduced to the form of Fig. 1, the equation for balance is

$$Z_{CD}Z_{AB} = Z_{BC}Z_{AD},$$

from which

$$Z_{CD} = \frac{Z_{BC}Z_{AD}}{Z_{AB}}. \tag{1}$$

Thus, if Z_{CD} is the unknown impedance, equation (1) evaluates it in terms of the other three impedances. Equation (1) is a vector equation and therefore the value of Z_{CD} both in magnitude and phase, or both components of it when considered as a complex quantity, may be obtained from this equation.

Although the above equations and subsequent discussion are based primarily on the use of impedances, it should be remembered that all of these relations may be obtained in the same general form if the bridge arms are considered as admittances.

The Bridge Requirements

If the impedances of equation (1) are replaced by the complex equivalents $R + jX$, then

$$R_{CD} + jX_{CD} = \frac{(R_{BC} + jX_{BC})(R_{AD} + jX_{AD})}{R_{AB} + jX_{AB}}. \tag{2}$$

From this equation R_{CD} and X_{CD} may be evaluated in terms of the other six quantities. Thus, if each component of the impedances of three arms is known, each component of the fourth impedance in terms of the other six components can be determined.

In obtaining the balance, any or all of the six component impedances occurring in the right hand side of equation (2) may be adjusted. Since there are two unknown quantities to be determined, at least two of these components must be adjusted. From the standpoint of simplicity and speed in operation and in order to keep the cost of the circuit to a minimum, it is desirable that not more than two of the known components be adjustable. It is also essential that the choice be such that a variation of one adjustable standard balance one component of the unknown, irrespective of the other component. In other words R_{CD} should be balanced by one known standard, this value of the standard being independent of the magnitude of X_{CD}, and, in turn, X_{CD} should be balanced by another standard, the value of which should be independent of the magnitude of R_{CD}. This condition of independent adjustment for the two components is essential for satisfactory operation of the bridge, since it allows the balance to be made more rapidly and systematically, and a given setting of one standard always corresponds to the same value of one component of the unknown, independent of the magnitude of the other component, thus allowing the calibration of each of the adjustable standards in terms of the unknown component which it measures.

To meet this requirement, the two components for use as adjustable standards should be so chosen that, when equation (2) is reduced to the general form

$$R_{CD} + jX_{CD} = A + jB, \tag{3}$$

where A and B are real quantities, one of the adjustable impedances will appear in A and not in B, while the other will appear in B but not in A.

Consideration of equation (2) shows that if adjustable standards consisting either of both components of Z_{BC} or of both components of Z_{AD}, are chosen, and if the impedances of the two remaining arms are selected so that their ratio is either real or imaginary, but not complex, then equation (2) reduces to the form of equation (3). No other combination will meet the requirement taking equation (2) as it stands. Since for the general case there is no essential difference in the resulting type of bridge whether Z_{AD} or Z_{BC} is used as our adjustable standard, this means that there is really only one method of adjustment, namely the use of both components of one adjacent impedance.

However, if it is realized that parallel components may be used instead of series components for the standard, then equation (2) may be rewritten as follows:

$$R_{CD} + jX_{CD} = (R_{AD} + jX_{AD})\,(R_{BC} + jX_{BC})(G_{AB} - jB_{AB}) \tag{4}$$

where

$$G_{AB} - jB_{AB} = Y_{AB} = \frac{1}{Z_{AB}}\cdot$$

From this it follows that G_{AB} and B_{AB} may be used as the adjustable standards, by making the product $Z_{AD}Z_{BC}$ real or imaginary.

Thus there are two methods of adjustment possible, either the two series components of an adjacent arm or the two parallel components of the opposite arm.

Having chosen the adjustable standards, there remain in each case two arms, adjacent in one case and opposite in the other, which have fixed values. These impedances must meet certain definite requirements, as already stated.

For the case of adjustment by an adjacent arm, that is, by Z_{AD}, equation (2) may be written in the form

$$R_{CD} + jX_{CD} = \frac{Z_{BC}}{Z_{AB}}(R_{AD} + jX_{AD}). \tag{5}$$

Then in order that this equation fulfill the requirements expressed by equation (3), the vector ratio of the fixed arms must be either real or imaginary but not complex, that is, the difference between their phase angles must be 0°, 180° or ± 90°.

For the case of adjustment by the opposite arm Z_{AB}, equation (4) may be written in the form

$$R_{CD} + jX_{CD} = Z_{BC}Z_{AD}(G_{AB} - jB_{AB}). \tag{6}$$

Then in order that this equation fulfill the requirements of equation (3), the vector product of the fixed arms must be either real or imaginary, but not complex, that is, the sum of their phase angles must be 0°, 180° or ± 90°.

In the case of bridges of the type indicated by equation (5), the fixed arms always enter the balance equation as a ratio, and are therefore called ratio arms, the bridges of this type being called ratio arm bridges.

In the case of bridges of the type indicated by equation (6), the fixed arms always enter the balance equation as a product, and are therefore called product arms, the bridges of this type being called product arm bridges.

These two types may be further subdivided according to whether the term involving the fixed arms is real or imaginary.

It should be pointed out at this time that the fixed arms are fixed in value only to the extent that they are not varied during the course of a measurement. They may be functions of frequency, and may be arbitrarily adjustable to vary the range of the bridge, but they are not adjusted in the course of balancing the bridge.

Classification of Bridge Types

The foregoing discussion shows that all simple four arm bridges meeting the requirements specified may be divided into four types. The balance equations of these four types may now be simply derived from the general equations (2) and (4).

1. *Ratio Arm Type—Ratio Real*

If Z_{BC}/Z_{AB} is real, then

$$\theta = \theta_{BC} - \theta_{AB} = 0° \text{ or } 180°.$$

That is

$$Z_{BC}/Z_{AB} = R_{BC}/R_{AB} = X_{BC}/X_{AB}. \tag{7}$$

Substituting equation (7) in equation (5) and separating,

$$R_{CD} = \frac{R_{AD}R_{BC}}{R_{AB}} = \frac{R_{AD}X_{BC}}{X_{AB}} \tag{8}$$

and

$$X_{CD} = \frac{X_{AD}R_{BC}}{R_{AB}} = \frac{X_{AD}X_{BC}}{X_{AB}}. \tag{9}$$

For this type it follows from equations (8) and (9) that the components of Z_{CD} are balanced by components of Z_{AD} of the same phase, that is R_{AD} will balance R_{CD}, and X_{AD} will balance X_{CD}.

2. *Ratio Arm Type—Ratio Imaginary*

If Z_{BC}/Z_{AB} is imaginary, then

$$\theta = \theta_{BC} - \theta_{AB} = \pm 90°.$$

That is

$$Z_{BC}/Z_{AB} = jX_{BC}/R_{AB} = -jR_{BC}/X_{AB}. \tag{10}$$

Substituting equation (10) in equation (5) and separating,

$$R_{CD} = -\frac{X_{AD}X_{BC}}{R_{AB}} = \frac{X_{AD}R_{BC}}{X_{AB}} \tag{11}$$

and

$$X_{CD} = \frac{R_{AD}X_{BC}}{R_{AB}} = -\frac{R_{AD}R_{BC}}{X_{AB}}. \tag{12}$$

For this type it follows from equations (11) and (12) that the components of Z_{CD} are balanced by components of Z_{AD} 90° out of phase, that is X_{AD} will balance R_{CD} and R_{AD} will balance X_{CD}.

3. *Product Arm Type—Product Real*

If $(Z_{BC}Z_{AD})$ is real, then

$$\theta = \theta_{BC} + \theta_{AD} = 0° \text{ or } 180°.$$

That is

$$Z_{BC}Z_{AD} = Z_{BC}/Y_{AD} = R_{BC}/G_{AD} = -X_{BC}/B_{AD}. \tag{13}$$

Substituting equation (13) in equation (6)

$$R_{CD} = \frac{G_{AB}R_{BC}}{G_{AD}} = -\frac{G_{AB}X_{BC}}{B_{AD}} \tag{14}$$

and

$$X_{CD} = -\frac{B_{AB}R_{BC}}{G_{AD}} = \frac{B_{AB}X_{BC}}{B_{AD}}. \tag{15}$$

For this type the components of Z_{CD} are balanced by components of Y_{AB} of the same phase, that is G_{AB} will balance R_{CD} and B_{AB} will balance X_{CD}.

4. *Product Arm Type—Product Imaginary*

If $(Z_{BC}Z_{AD})$ is imaginary, then

$$\theta = \theta_{BC} + \theta_{AD} = \pm 90°.$$

That is

$$Z_{BC}Z_{AD} = Z_{BC}/Y_{AD} = jR_{BC}/B_{AD} = jX_{BC}/G_{AD}. \tag{16}$$

Substituting equation (16) in equation (6)

$$R_{CD} = \frac{B_{AB}R_{BC}}{B_{AD}} = \frac{B_{AB}X_{BC}}{G_{AD}} \tag{17}$$

and

$$X_{CD} = \frac{G_{AB}R_{BC}}{B_{AD}} = \frac{G_{AB}X_{BC}}{G_{AD}}. \tag{18}$$

For this type the components of Z_{CD} are balanced by components of Y_{AB} 90° out of phase, that is B_{AB} will balance R_{CD} and G_{AB} will balance X_{CD}.

The relations given in these equations are summarized in Table I.

TABLE I

BRIDGE TYPES

Unknown	Adjustable Standard			
	Ratio Arm Type		Product Arm Type	
	Ratio Real	Ratio Imaginary	Product Real	Product Imaginary
R_{CD}	R_{AD}	X_{AD}	G_{AB}	B_{AB}
X_{CD}	X_{AD}	R_{AD}	B_{AB}	G_{AB}
G_{CD} [1]	G_{AD}	B_{AD}	R_{AB}	X_{AB}
B_{CD} [1]	B_{AD}	G_{AD}	X_{AB}	R_{AB}

[1] These values may be derived by using admittances in place of impedances and vice versa throughout.

ACTUAL BRIDGE FORMS

The fixed arms may be made up of simple resistances or reactances or of complex impedances provided they meet their phase requirements. Since the choice of complex impedances has no practical advantages over simple reactances or resistances, the choice of fixed impedances should obviously be made on the basis of the simplest practical type. So they will be limited for the present to simple resistance, capacitance, and self inductance.

Fig. 2 gives all of the combinations of fixed arms which meet the phase angle requirements already stated, when limited to simple resistance, inductance, or capacitance. For all forms, the magnitude of one arm is given in terms of the other and of a constant K, such

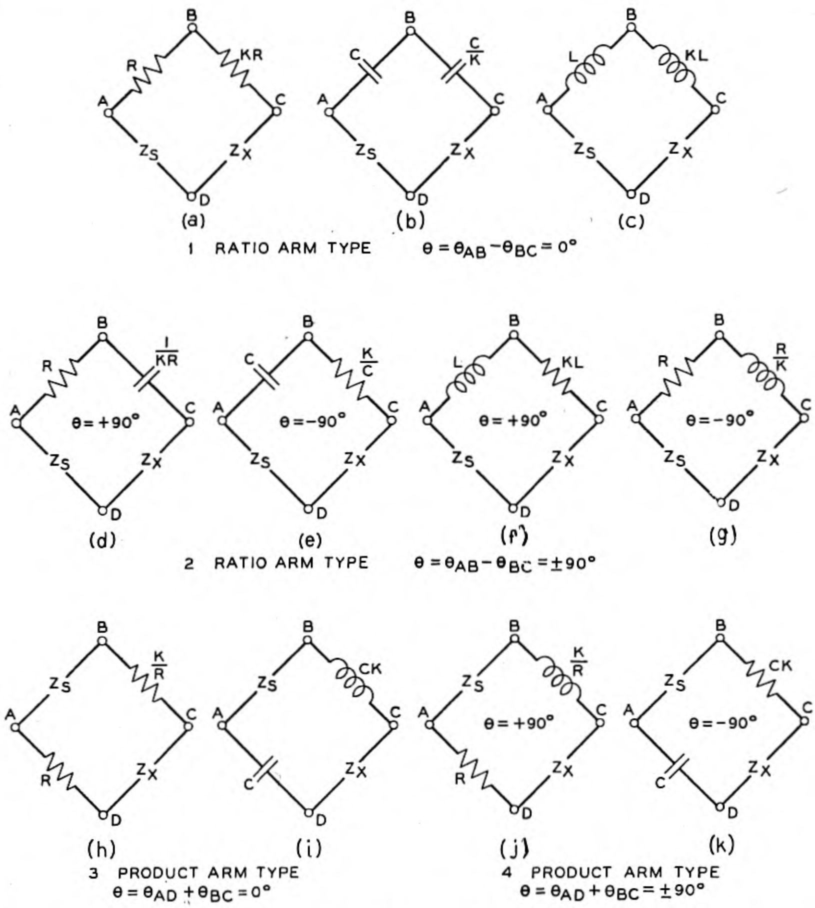


Fig. 2—The various forms of 4-arm bridges divided into four types. Forms f, g and j are impractical.

that the only term which appears in the balance equation is the term K. None of these bridges represents a distinctly new type, but since the classification is by means of the fixed impedance arms, one of them may be used to measure several types of impedance. Accordingly, it may correspond to more than one of the well-known bridge types.

For this reason, any references to, or comparison with existing special types of bridge are omitted.

TABLE II

BALANCE EQUATIONS

Unknown	Ratio Arm Type			Product Arm Type		
	$\theta = 0$	$\theta = +90°$	$\theta = -90°$	$\theta = 0$	$\theta = +90°$	$\theta = -90°$
$R_{CD} =$	KR_{AD}	KL_{AD}	K/C_{AD}	KG_{AB}	K/L'_{AB}	KC'_{AB}
$L_{CD} =$	KL_{AD}	—	KR_{AD}	KC'_{AB}	KG_{AB}	—
$C_{CD} =$	KC_{AD}	$1/KR_{AD}$	—	KL'_{AB}	—	$1/KG_{AB}$
$G_{CD} =$	KG_{AD}	$1/KL'_{AD}$	C'_{AD}/K	R_{AB}/K	L_{AB}/K	$1/KC_{AB}$
$L'_{CD} =$	KL'_{AD}	—	K/G_{AD}	C_{AB}/K	K/R_{AB}	—
$C'_{CD} =$	KC'_{AD}	G_{AD}/K	—	L_{AB}/K	—	R_{AB}/K
Figures....	2A 2B 2C	2D 2F [2]	2E 2G [2]	2H 2I	2J [2]	2K

[2] These forms are not practical.

R, L and C = series components of complex arms.

G, L' and C' = parallel components of complex arms.

K has the value indicated on the individual circuits of Fig. 2.

$$\theta = \theta_{AB} - \theta_{BC} \quad \text{for Ratio Arm Type}$$

$$\theta = \theta_{AD} + \theta_{BC} \quad \text{for Product Arm Type}$$

Table II gives the balance equations for each type of bridge for the measurement of any component of the unknown impedance in terms of resistance, capacitance, and inductance. These equations are simply derived from the general equations (8) to (18) by substitution of circuit constants for impedances and by the introduction of the constant K. This constant must be evaluated from the relation between the ratio arms or product arms shown in the individual bridge forms of Fig. 2. At the bottom of Table II are given the corresponding bridge figures for reference. This table shows no bridges having a phase relation of 180° between the fixed arms. A little consideration will show that since the phase relation between the unknown and the standard for such bridges must also be 180°, they cannot be used to measure any but pure reactances or negative resistances. Accordingly, they are not considered herein. In the case of the 90° relation, both signs must be considered and result in bridges which are complimentary with respect to one another, that is while one measures only inductive impedances, the other measures only capacitive impedances. Thus

Table II shows the imaginary type subdivided into two subtypes, depending on the sign of the angle.

As an example of the use of this table: Suppose it is desired to measure the series resistance and inductance of an unknown impedance. This may be done by using adjustable standards of series resistance and inductance, series resistance and capacitance, parallel resistance and capacitance, or parallel resistance and inductance, by choosing the particular type of bridge for the purpose. For instance, referring to Table II, if it is desired to measure the series resistance in terms of conductance, and the series inductance in terms of parallel capacitance, the product arm bridge with real ratio, that is either Fig. 2*h* or 2*i*, would be used.

Since there are six types of balance equations given in Table II, it follows that five of the circuits of Fig. 2 are duplicates of others from the standpoint of the balance equations which they give. For instance, there is no difference whatever in the theoretical operation of the bridges of Figs. 2*a*, 2*b*, and 2*c*. The choice must be determined entirely from other considerations. In the same way, as indicated by the figures tabulated in Table II, Figs. 2*d* and 2*f* give identical results as do Figs. 2*e* and 2*g*, and Figs. 2*h* and 2*i*. From the practical standpoint, there may be, and actually there is, considerable difference in the merits of these different forms. At this time, we may simply state that where a choice is possible, resistance is the preferred form of fixed arm and capacitance is preferred to inductance. This allows us to choose our preferred forms as Fig. 2*a*, Fig. 2*d*, Fig. 2*e*, and Fig. 2*h*.

A study of Table II shows that bridges of fixed ratio arm type always measure the series components of the unknown in terms of series components of the standard and, conversely, they measure the parallel components in terms of parallel components of the standard. Bridges of product arm type measure the series component of the unknown in terms of parallel components of the standard and conversely.

None of the balance equations of Table II includes frequency, that is, all of them allow the evaluation of each component of the unknown directly in terms of a corresponding component of the standard with the exception that in some cases the relation is a reciprocal one. Practically any form of standard may be chosen in order to measure a given type of unknown impedance.

Practical Considerations

So far the question whether the requirements for the fixed arm impedances given in Fig. 2 can be met in practice has not been considered. It may be well to point out that the performance of the bridge is determined very much by the degree to which the phase angle requirements are met. If there is appreciable error here, the two balances will not be entirely independent and necessary corrections will be complicated and difficult to make. Consequently, the first essential for a satisfactory bridge is that its fixed arms meet their phase angle requirements. For a general purpose bridge these requirements must hold independent of frequency at least over an appreciable frequency range.

The forms given in Fig. 2 meet their phase angle requirements at all frequencies provided the arms are actually pure resistances or reactances. If they have residuals associated with them, it is still possible to meet the phase angle requirements in most cases, at least over a reasonable frequency range, as discussed below.

Resistances can be made to have practically zero phase angle, and condensers, particularly air condensers, may be made to have phase angles of practically 90°. In the case of condensers having dielectric loss, this loss may be kept quite small. However, it takes such a form that the phase difference of the condenser is approximately independent of frequency. For this reason, it can not be represented accurately either as a fixed resistance in series with the condenser or as a fixed conductance in shunt, when considered over a frequency range. Due to the small amount of this loss, it is usually satisfactory to represent it in either one form or the other, whichever is the more convenient.

In the case of inductance, there is always a quite appreciable series resistance which, for the usual size of coil, can not be neglected and must accordingly be corrected for.

With the above considerations in mind, the forms of Fig. 2 may now be reconsidered from the practical standpoint. It is readily seen that the requirements of the real ratio type bridge can be met using resistances, capacitances, or inductances. In the case of the imaginary ratio type, the requirements can be met, at least very approximately, in the case of Figs. 2*d* and 2*e*. However, in the case of Figs. 2*f* and 2*g*, any resistance in series with the inductance must be corrected by a capacitance in series with the resistance, if the correction is to be independent of frequency. Since the value of this series capacitance will, in general, be large, this form of correction is unsatisfactory. For instance, for a bridge in which the value of R is 1000 ohms and the inductance has a high time constant, the series capacitance required is in the order of 3 μf. By using a standard of inductance having larger series resistance, we may reduce this capacitance, but we then have a form of bridge which is, in effect, a compromise between Figs. 2*f* and 2*g*, and Figs. 2*d* and 2*e*, which has no practical advantages over the latter. Accordingly, the forms of Figs. 2*f* and 2*g* must be considered impractical, particularly as Figs. 2*d* and 2*e* give identical performance.

In the case of the product arm type the requirements can be met by Fig. 2*h* and can be met by Fig. 2*i* by adding a conductance in shunt with the capacitance to compensate for the series resistance of the inductance. However, even though this allows us to meet the requirement, this form is less satisfactory than that of Fig. 2*h* due to the difficulty of designing an inductance standard having inductance and series resistance invariable over an appreciable frequency range. Again the requirements can be readily met by Fig. 2*k*, but in the case of Fig. 2*j* series resistance of the inductance can be corrected only by shunting the resistance arm by pure inductance, which is impractical. This is unfortunate since it rules out one form of bridge for which there is no duplicate and, consequently, makes the measurement of inductive impedances by bridges of this type impractical.

Summarizing the above, practical considerations rule out Figs. 2*f*, 2*g*, and 2*j*, reducing to five the number of different bridge types. There are eight forms remaining, namely three of the real ratio type, each capable of giving the same performance; two of the imaginary ratio type which are complementary, together giving a measurement of inductive and capacitive impedances; two of the real product type which will measure all types of impedance; and one imaginary product type which is capable of measuring only capacitive impedances.

The only duplicate forms are in the case of the real ratio and real product types. In the case of the latter, Fig. 2*h* is to be preferred in practically all cases to Fig. 2*i*, as already explained, and thus we can say that, practically speaking, we have duplicate forms only in the case of the real ratio type.

The three forms of this type are all used and each has certain advantages for certain types of measurements. This type of bridge, commonly known as the direct comparison type, is probably used more than any other, and is one of the most accurate types, particularly in the special case of equal ratio arms. This is due to the fact that a check for equality of the ratio arms may be readily made by a method of simple reversal without any external measurements, and by this means practically all the errors of the bridge may be eliminated. Resistance ratio arms are preferable for a general purpose bridge because they are more readily available and more readily adjusted to meet their requirements. They also give an impedance independent of frequency, which is usually desirable. Capacitance ratio arms have certain advantages for particular cases. They may be readily chosen to give high impedance values, this being an advantage in certain cases, for instance in the measurement of small capacitances at low frequencies. This form is also desirable where high voltages must be used, since the ratio arms may be designed to withstand high voltages without the dissipation of appreciable energy. It also has the advantage that where measurements are desired with a direct current superimposed on the alternating current, the direct current is automatically excluded from the ratio arms and thus all of the direct current applied to the bridge passes through the unknown and there is no dissipation due to the direct current in the ratio arms. The impedance of the ratio arms decreases as the frequency increases, which is usually a disadvantage but may have advantages in some cases, such as the measurement of capacitance. There may be a disadvantage, in some cases, due to the load on the generator being capacitive, thus tending to increase the magnitude of the harmonics, and again, in the case of the measurement of inductances, there may be undesirable resonance effects.

The inductance ratio arm type has advantages where heavy currents must be passed through the bridge, since the ratio arms of this type may be designed to carry large currents with low dissipation. A modification of this type, where there is mutual inductance between the ratio arms, gives the advantage of ratio arms of high impedance with a corresponding low impedance input. A further modification consists in making the ratio arms the secondary of the input transformer, thus combining in one coil the functions of ratio arms and input transformer. This form, of course, departs from the simple four-arm bridge, but is mentioned here due to its simplicity and actual practical advantages.

Substitution Methods

In any of the bridges discussed and, in fact, in practically all bridges, it is possible to evaluate the unknown by first obtaining a balance with the unknown in the circuit and then substituting for it adjustable standards which may be adjusted to rebalance the bridge. This is, in general, a very accurate method, eliminating to a large degree the necessity for the bridge to meet its phase angle requirement. However, in the case of complete substitution of standards to balance both components of the unknown, the method has no advantage except accuracy over the bridges of type 1, Fig. 2, since standards of the same type as the unknown must be used and, in general, this method lacks the flexibility of bridges of type 1, obtained by their unequal ratio arms. On the other hand, the use of substitution to measure the resistance or conductance component of the unknown has many advantages, the principal one being that it allows the choice of a type of bridge which will give directly the reactance component of the unknown in terms of an adjustable resistance and then by use of the substitution method to balance the resistance or conductance of the unknown by means of a second adjustable resistance, thus obtaining the ideal method of balance, using two adjustable resistances.

For the purpose of illustration, the case of the measurement of an inductive impedance may be taken. In general, the most desirable method would be to balance the reactance by means of series resistance. This can be done by means of the bridges of Figs. 2*e* or 2*g*. Choosing Fig. 2*e* as the preferred form, the bridge would normally take the form of Fig. 3*a*.

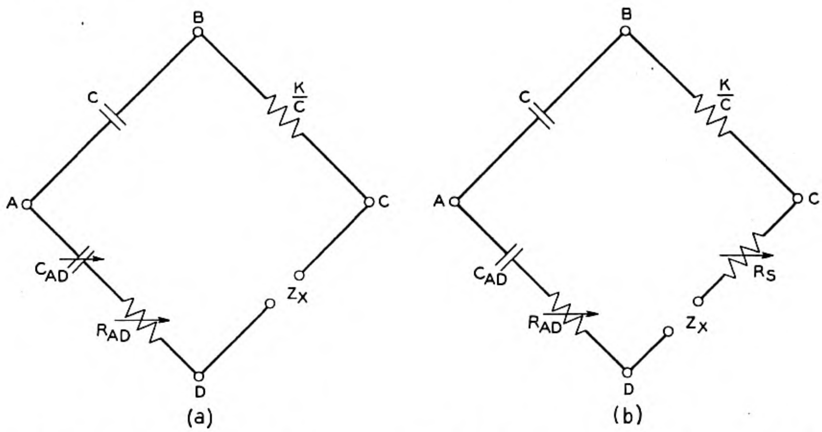


Fig. 3—(*a*) Bridge of type 2 for measuring self-inductance. (*b*) The same bridge modified by the use of partial substitution.

For normal operation, C_{AD} and R_{AD} would be the adjustable standards. The series inductance of the unknown would be given directly as KR_{AD}, while the series resistance would be given as K/C_{AD}. This measurement of the series resistance requires an adjustable capacitance and a computation due to the reciprocal relation. Now suppose a fixed value for C_{AD} were used and an adjustable resistance standard R_S placed in series with Z_X, giving the form of Fig. 3*b*, in which R_{AD} and R_S are the adjustable standards. If terminals Z_X are short circuited, the conditions for balance are $R_S = K/C_{AD}$ and

$R_{AD} = 0$. Then the unknown Z_X is inserted and the bridge rebalanced. The inductance of the unknown is given, as for Fig. 3*a*, as KR_{AD}, but since C_{AD} is unchanged the total resistance in CD is unchanged. Therefore, the series resistance of the unknown will be equal to the change in R_S between the two balances.

This bridge circuit may be recognized as the familiar bridge due to Owen,[3] and it is, theoretically at least, when used as described, an exceedingly desirable bridge for inductance measurements.

It should be pointed out here that since either C_{AD} or R_S may equally well be used to balance R_X, it is not necessary to use either one or the other exclusively in any one bridge. The adjustments may be combined so that the capacitance adjustment will take care of large changes and R_S of small changes; that is, C_{AD} may be used for coarse adjustment and R_S for fine adjustment. This compromise is, in general, more satisfactory than either method used alone.

The imaginary product arm type, particularly the form of Fig. 2*k*, is also well adapted to modification to enable it to measure capacitance and conductance in terms of two adjustable resistances.

There is a further modification of the substitution method, which is in common use. As already explained, there is little practical advantage in the substitution method for measuring either inductance or capacitance. However, there are occasions where the substitution of capacitance for inductance has advantages. Since the reactance of one is opposite in sign to that of the other, the method might more correctly be termed a compensation method, but in common with other substitution methods it can be made irrespective of the type of bridge. Various modifications of the general method may be used, but they are all classed under the general head of resonance methods.

Resonance Methods

If it is desired to measure the inductance of any inductive impedance, a capacitance standard may be inserted in series with it, and adjusted until the total reactance of the combination is zero. The only function the bridge performs is to measure the effective resistance of the combination and to determine the condition of zero reactance. Any of the bridges of Fig. 2 will do this satisfactorily, but those of real ratio type, that is the simple comparison type, are the most satisfactory since they give the resistance directly in terms of an adjustable resistance standard. This type of bridge is usually termed a series resonance bridge. The value of the inductance is computed from the resonance formula $\omega^2 LC = 1$. It has the dis-

[3] D. Owen, *Proc. Phys. Soc.*, London, October, 1914.

advantage that it involves the frequency, but it has the compensating advantage that the method, being essentially a direct measurement of the resistance of the resonant circuit, is very accurate for the measurement of effective resistance.

The condenser may equally well be shunted across the unknown, in which case the bridge circuit is called a parallel resonance bridge. However, if the ratio of reactance to resistance of the unknown is not high, the expression for the series inductance in this case is not as simple as that for series resonance, and is not independent of the value of the effective resistance, that is the two adjustments are not independent.

Fig. 4 shows the forms taken by the *CD* arm for resonance measurements. Fig. 4*a* is the series resonant circuit using an adjustable capacitance standard. Fig. 4*b* is the parallel circuit using an adjustable capacitance standard.

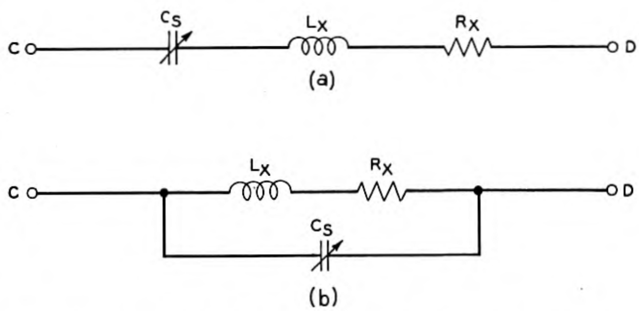


Fig. 4—(*a*) The CD arm of the bridge as used for series resonance measurements. (*b*) The CD arm of the bridge as used for parallel resonance measurements.

Some Theoretical and Practical Aspects of Noise Induction*

By R. F. DAVIS and H. R. HUNTLEY

This article discusses the physical processes of induction between neighboring power and telephone lines and describes means by which certain phenomena of interest in this connection have been qualitatively demonstrated to power and telephone employees.

INTRODUCTION

EARLY in the development of the power and telephone industries, serious problems were encountered because of induction between neighboring power and telephone circuits. In 1885, about 150 representatives of Electric Light Companies assembled in Chicago and discussed the many problems of interference with telephone service due to induction which were even then coming up. This meeting resulted in the formation of the National Electric Light Association.

Prior to this time all telephone circuits were grounded, that is, they used a single wire with ground return, and so were very susceptive to inductive disturbances. There was also a great deal of interference between different telephone circuits on the same line (that is, cross-talk) so that conversations on one circuit could be overheard on others. General John J. Carty, then working in Boston, had been doing a great deal of work on this subject and by about the end of 1885 had not only developed the metallic telephone circuit, which employs two wires and does not use the earth as part of the circuit, but also had worked out methods of applying transpositions. These developments afforded such a large reduction in the susceptiveness of the circuits to external influences that the problems of coordination existing at that time were largely solved.

However, with the expansion and development of the power and telephone industries, new problems of coordination arose, and the nature and control of the phenomena involved have been the subject of continuous study by both industries. While a great deal has been learned about the technical phases of the problem and the best methods of handling it, the coordination of the plants of power and telephone companies in such a way that safety and service are promoted with minimum expense still involves important problems. These problems not only concern the engineers who are responsible for plant design and for technical advice, but also enter into the work of the field forces who

* This paper appeared in somewhat different form in *Amer. Railway Assoc. Proc.*, June, 1932, under the title "Demonstration and Talk on Noise Induction" by H. R. Huntley.

actually construct, operate and maintain the plants and into the considerations of management. Naturally, the best results can be secured if all concerned have a thorough understanding of the subject and appreciate each other's requirements and points of view.

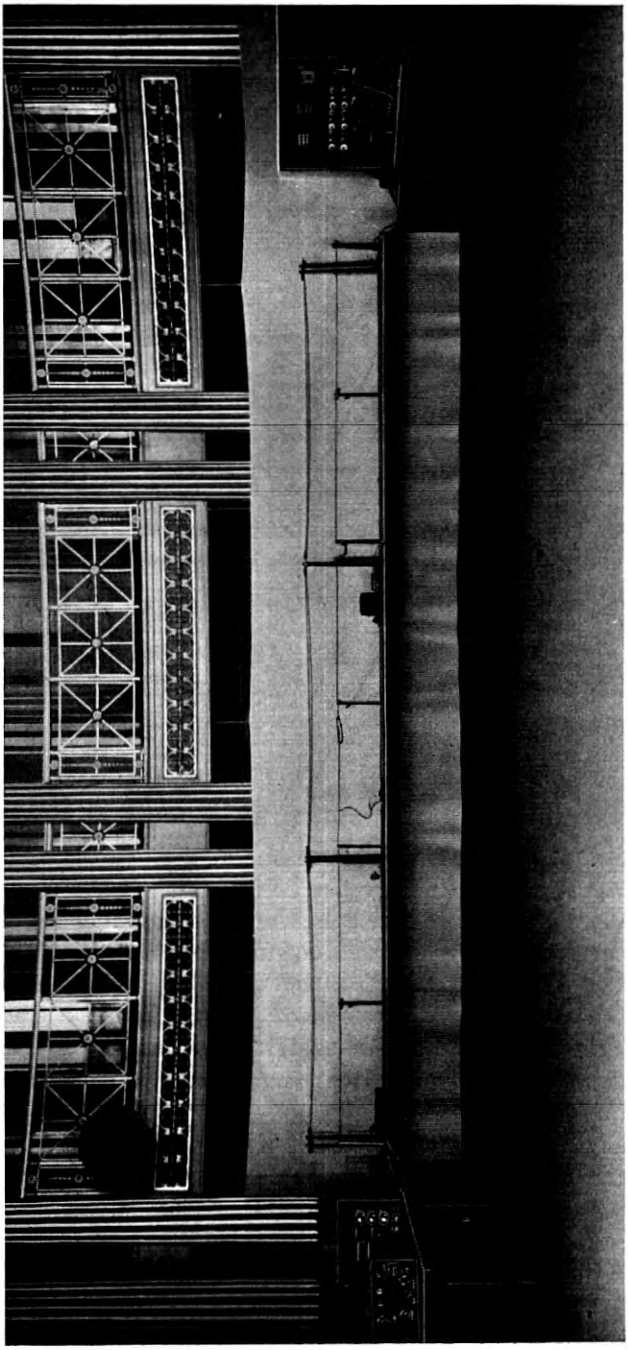
In promoting the mutual understanding of this subject which is so desirable, it has been found helpful in some cases to use demonstrations of the principles underlying the work accompanied by explanations in everyday language. One of the demonstrations which has been shown before a number of audiences of power and telephone people with this in mind has to do with noise frequency induction and employs the miniature lines and apparatus shown in photograph No. 1. A considerable amount of interest has been aroused by these demonstrations and many of the people in the audiences have found complete or partial explanations of some specific problems which have been troubling them.

In order to illustrate the manner in which the miniature lines and apparatus may be used to demonstrate principles of noise frequency induction, there follows a description of this apparatus and a discussion of the processes of induction along the lines usually followed in the demonstrations.

Fundamentals of Problem

The problems concerned with inductive coordination arise due to the fact that wires transmitting electricity necessarily have electric and magnetic fields about them which may under certain conditions cause voltages to appear in other wires which are in these fields. This phenomenon is called induction. The voltages and currents used in power transmission are much greater than those used in speech transmission so that there are practically no situations in which the currents and voltages on telephone systems affect power system operation due to induction, but situations do arise in which power system voltages and currents affect telephone system operation.

The effects of induction in a given situation of proximity between power and telephone circuits are dependent upon the characteristics of both the power and telephone systems and upon the coupling (due to the electric and magnetic fields) between them. It is theoretically possible for a power line to be so constructed and maintained that it would cause no induction into a nearby telephone circuit. Such a power line would be said to have zero "inductive influence." Likewise, it is theoretically possible to have a telephone circuit so constructed and maintained that it would be unaffected by any electric or magnetic fields set up by power systems. Such a telephone circuit would be said to have zero "inductive susceptiveness." Also, of



Photograph No. 1.

course, regardless of the characteristics of the power and telephone circuits, if the separation between them could be very great, there would be no "inductive coupling" and consequently, no induction from one into the other. Practically, of course, neither power nor telephone systems can be constructed so as to have zero influence or susceptiveness, and it is frequently impracticable to separate them sufficiently to make the coupling negligible. The practical coordination problem, therefore, is to work out the most convenient and economical method of controlling the factors so that inductive interference is avoided.

In the practical problem of inductive coordination between power and telephone systems there are often two more or less distinct aspects to be considered. One of these aspects is concerned with the possibility of extraneous currents in the telephone circuits which have frequencies within the range used in transmitting speech and which may, therefore, cause "noise" in the telephone receivers at the ends of the circuit. This phenomenon may arise during the normal operation of power and telephone systems although abnormal conditions on either system may result in increasing the noise during the existence of such abnormal conditions. The other aspect commonly referred to as "low frequency induction," is associated almost entirely with faults to ground on power systems and is primarily concerned with the possibility at such times of high induced voltages at fundamental power system frequency. This article, however, is confined to the noise aspect of the problem.

Demonstration Apparatus

In order to qualitatively illustrate some of the factors involved in noise induction, a miniature inductive exposure as shown in the photograph referred to previously, may be used. The demonstration circuits consist essentially of a miniature three-phase, three-wire power line and a two-wire telephone line which are set parallel to each other on a grounded copper screen and are connected as shown schematically in Fig. 1. The power line can be energized in various ways from an ordinary three-phase power distribution circuit through suitable transformers. The telephone line is connected to an amplifier and loud speaker so that the noise on the telephone circuit under various conditions can be heard. Both lines can be transposed independently or in a coordinated manner and unbalances can be inserted in the telephone circuit. The particular connections and arrangements of the lines and apparatus used in each of the demonstrations are described as that demonstration is discussed.

With an inductive exposure of the limited dimensions available, it is impracticable to secure results which can be related in a quantitative sense to field conditions. Also, such effects as the shielding between

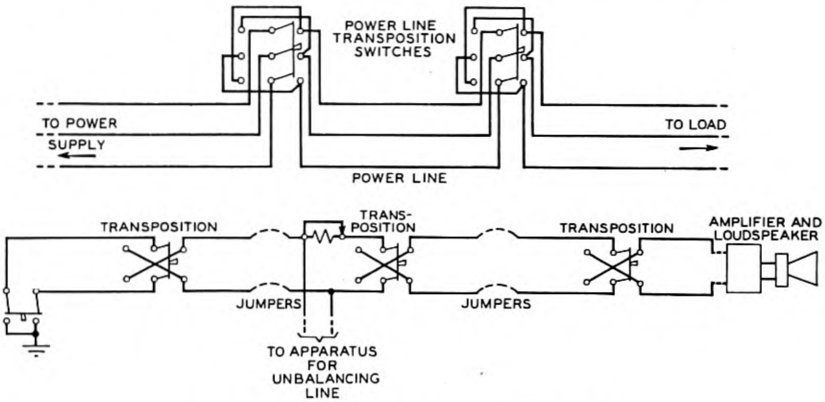


Fig. 1—Schematic of demonstration circuit.

the various telephone circuits on a multi-wire line, propagation effects, etc., cannot be shown. Furthermore, the exposure is a great deal more regular than those usually encountered in practice so that, for example, a higher effectiveness of transpositions than is usual can be secured. However, many of the fundamentals of the problem can be illustrated qualitatively.

Nature of Magnetic and Electric Induction

It is often desirable to consider effects of magnetic and electric induction separately, particularly in the technical analyses of specific problems. This is not only because the physical processes and the effects of voltage and current induction are quite different but also because the power circuit voltages and currents are often affected differently by changes in conditions. "Electric induction" is a term used to refer to induction due to the voltages on the power line, while "magnetic induction" is used in connection with the inductive effects of currents.

Considering electric induction first, perhaps the simplest method of visualizing the phenomenon, is by means of the capacitances involved with a single power wire and a single telephone wire as shown in Fig. 2. Neglecting the impedances outside the exposure (which are shown dotted in Fig. 2) the voltage of the power wire to ground (E_P) divides over the capacitances C_{TP} and C_{TO} in proportion to their impedances (that is, in inverse ratio to their capacitance values). The induced voltage on the telephone wire may therefore be expressed mathematically as:

$$E_T = \frac{C_{TP}}{C_{TO} + C_{TP}} E_P.$$

Where there are numerous power and telephone wires, capacitances exist between every possible combination of wires, and of wires and ground, resulting in a complicated network, but the principles involved are the same as in the simple case discussed above.

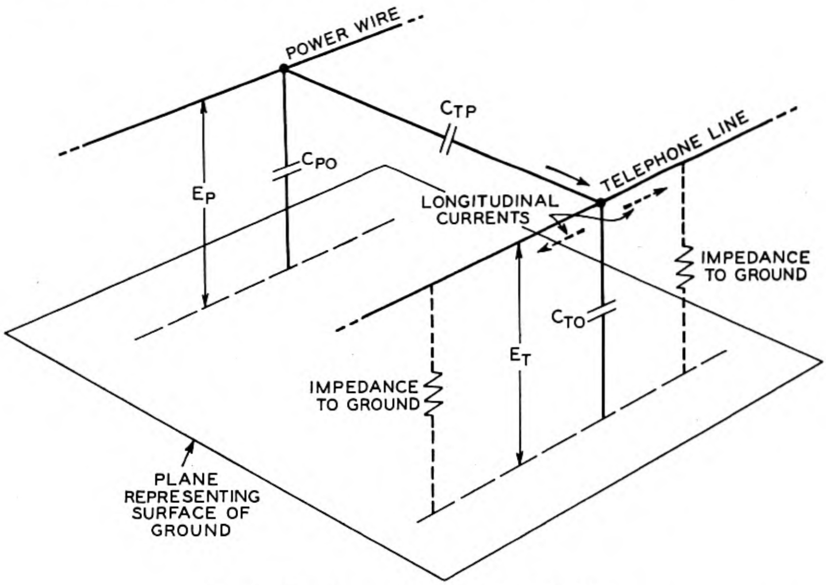


Fig. 2—Fundamental of electric induction.

The point of particular interest is that the potential of the telephone wire tends to be the same all along its length and, if it is perfectly insulated from ground, extends only through the length of the exposure, and has no equipment on it, this potential is independent of the length of the exposure (this is the condition shown in Fig. 2 if the impedances to ground are neglected). This is because, while all of the capacitances in the above equation are proportional to exposure length, the *ratio* $\frac{C_{TP}}{C_{TO} + C_{TP}}$ is independent of length. However, in the usual field case, the circuits extend beyond the exposure and have equipment connected between them and ground so that there are impedances to ground outside the exposure (as shown dotted in Fig. 2) through which longitudinal currents will flow. The net voltage to ground under these conditions is equal to the total of the longitudinal currents in the two directions times the impedances to ground looking in the two directions considered in parallel and, since these impedances are usually much smaller than the impedance through which the current reaches the telephone line (capacitance C_{TP}), this voltage is usually much smaller than the *induced* voltage (see equation above). Since the impedance of C_{TP} controls the total longitudinal current, this current will be practically independent of the telephone circuit impedances to ground and will be proportional to exposure length. It will also be proportional to the frequency of the harmonics in the inducing voltage (since the impedance of a capacitance is inversely proportional to frequency). Hence, for given telephone circuit impedance conditions (outside the exposure) the voltage to ground will be proportional to exposure length and to the frequency of the inducing harmonics in a uniform (electrically short) exposure.

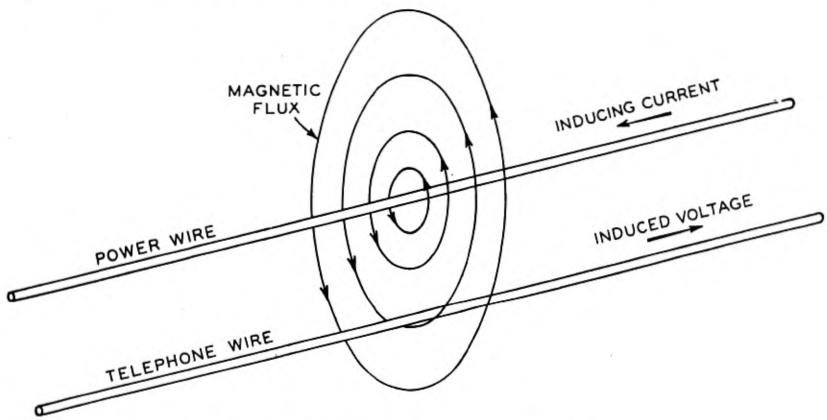


Fig. 3—Fundamental of magnetic induction.

Considering magnetic induction, the current in the power wire sets up a magnetic field which alternates at the frequency of the current. If a telephone wire is located in this field, a voltage is induced *along* it which is proportional to the rate of change of the magnetic flux just as a winding in a transformer has a voltage induced along it. This phenomenon is illustrated in Fig. 3. The voltage between the telephone circuit and ground varies from point to point along the circuit and depends on the distribution of the impedances to ground as well as on the distribution of the induced voltage. Also since the voltage acts along the circuit and the part induced in each short length adds directly to those in all other short lengths, the total induced voltage is directly proportional to the exposure length in a uniform (electrically short) exposure. Also, since the rate of change of magnetic flux is proportioned to frequency, the induced voltage will be proportional to the frequency of the harmonics in the inducing current.

The demonstration which shows the fundamental difference in the action of electric and magnetic induction is shown in Fig. 4.

1. In Fig. 4-*A* the arrangements for demonstrating electric induction as well as the way the induced voltage acts through the impedance to earth in the exposure are shown. In the setup the power line is energized at about 200 volts, balanced 3-phase, but since the far end is open the current in it is negligible. Consequently only electric induction is present in appreciable amount. Since the voltage to ground of the telephone circuit is the same over its entire length, grounding it at any point reduces the voltage at all points. This is shown in the demonstration by the great reduction in the noise to ground as heard in the loud speaker when the switch at the far end of the line is closed thus grounding the line.
2. In Fig. 4-*B* the arrangements for demonstrating magnetic induction as well as the manner in which the induced voltage acts are shown. In this setup the power line is energized at about 17 volts, 3-phase and has a load such that the current is about 15 amperes in each wire. Due to the low voltage and the relatively large current, magnetic induction is predominant. Since the induced voltage acts *along* the circuit, it can be prevented from acting on the amplifier input by opening the circuit at any point. This is indicated in the demonstration by the fact that the noise in the loud speaker is much greater when the switch at the far end of the line is closed than when it is open. (This is, of course, the exact reverse of the conditions when electric induction was being demonstrated.)

In the demonstrations the lines used are very short electrically. For circuits which are long enough so that propagation effects must be considered, the results of grounding or opening the far end of the circuit may be considerably different than for electrically short circuits.

Inductive Coupling

General

In discussing inductive coupling, it is necessary to consider not only the metallic power circuit and the metallic telephone circuit but also the

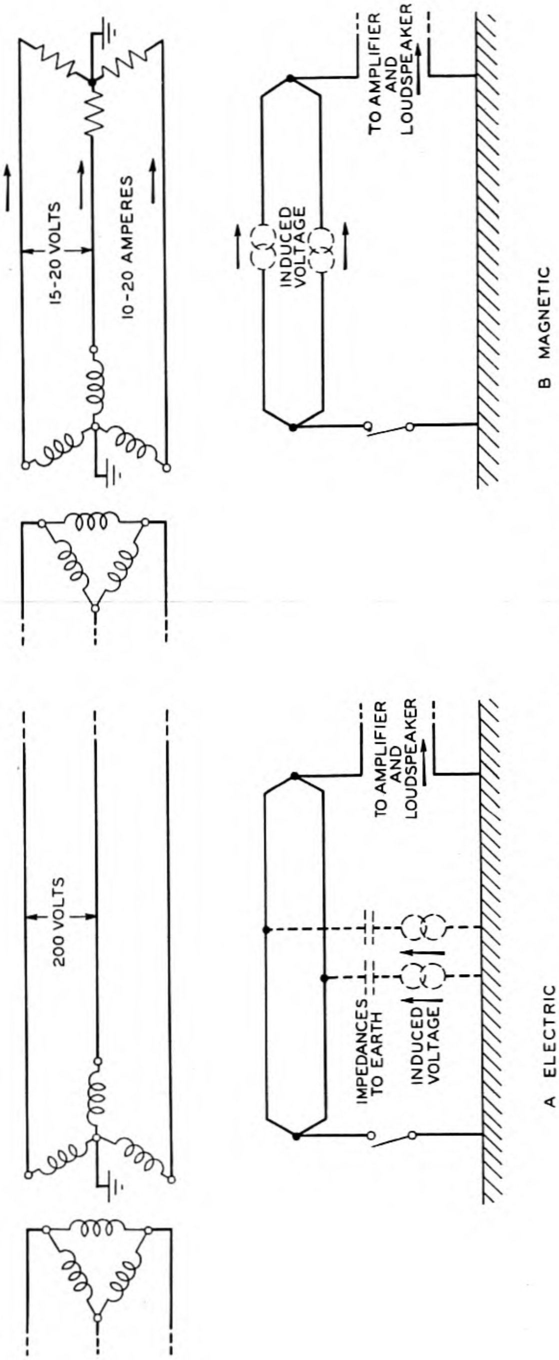


Fig. 4—Demonstration of electric and magnetic induction.

circuit composed of the power wires in parallel with ground return and the circuit composed of the telephone conductors in parallel with ground return. This is because, while the power to customers is usually transmitted over metallic power circuits and telephone conversations between telephone customers are usually over metallic telephone circuits, the circuits composed of the wires and ground in both systems enter into the induction picture unless the systems are perfectly balanced (which, as pointed out previously, is impracticable).

Considering the power system first, it is customary to divide the line currents and voltages into residual and balanced components. The balanced currents are the components which add up vectorially to zero. The residual current is the vector sum of the line currents and is that which remains after the balanced components are taken out. Similarly, the balanced voltages are the components of the voltages to ground which add up vectorially to zero and the residual voltage is the vector sum of the voltages to ground.

Thus it is seen that the balanced voltages and balanced currents are confined to the line wires while the residuals act in the circuit composed of the line wires in parallel with earth return. For a three-phase circuit the effect is that of a single-phase voltage equal to one third the residual voltage applied between the line wires and earth and a single-phase current equal to the residual current flowing out in the three phase wires in parallel and returning via the earth (or metallic paths other than the phase wires if such exist).

Whether appreciable residuals exist on the power system depends on many conditions, some of which are discussed later.

Considering the telephone circuit, the voltages, as pointed out in connection with the discussion of the theory of magnetic and electric induction, exist along the conductors or between them and earth. However, these voltages may not be identical for the two conductors of a metallic circuit and the vector difference exists as a voltage acting between the two wires. This voltage which, of course, tends to send current around the metallic circuit (and hence noise in the receivers at the ends of the circuit), is often spoken of as due to "direct metallic-circuit induction." The average of the voltages between the two wires and earth is often spoken of as "voltage to ground" and the currents in the two wires in parallel are often spoken of as "longitudinal-circuit" currents. The effects of these voltages to ground and longitudinal-circuit currents on telephone circuits which are not perfectly balanced are discussed later.

All of the factors which have been mentioned, that is, balanced and residual components, direct metallic induction, longitudinal circuit currents, etc., enter into the consideration of coupling. It is, of course, impracticable to do more in this discussion than consider some of the more important aspects of this phase of the subject.

In general, it can be said that except for very small separations where rapid changes in coupling may occur with changes in the relative positions of the circuits, all of the types of coupling will become smaller as the separation between the power and telephone circuits increases. The rate at which the coupling falls off with increasing separation depends on many factors. For example, the coupling involved in direct metallic induction generally falls off faster with increasing separation than does the coupling affecting the longitudinal telephone circuit. Likewise the coupling affecting the induction from balanced currents and voltages generally falls off faster than that from residual currents and voltages.

In order to demonstrate that, in general, the coupling is reduced by increasing the separation, the telephone line in the exposure is moved in such a way as to change the separation and it is noted that, as the separation increases, the noise decreases and vice versa.

For a uniform exposure, the amount of noise in an untransposed telephone circuit exposed to an untransposed power circuit will generally be approximately proportional to the length of the exposure, provided the total exposure is electrically short. (For long exposures, this proportionality may not hold because of phase-shift, attenuation effects, etc.) In order to illustrate the effect of changes in length of exposure, one-third, two-thirds, and all of the telephone line in the miniature exposure are employed successively and it is noted that the volume of sound from the loud speaker is approximately proportional to the length of the exposure. The direct proportionality between noise and exposure length does not hold for exposures to which coordinated transposition layouts have been applied as the resultant noise in such cases depends largely on the effectiveness of the coordinated layout. The effects of transposition are discussed in the following.

Transpositions in Power Circuits

Transpositions in power circuits are used primarily to accomplish two results. The first of these is the reduction, within exposures, of the induction from balanced currents and voltages. The second is the equalization of the admittances to earth and the series impedances of the power wires in order to limit the residual voltages and currents. In this discussion only the first of these two results (that is, reduction of induction due to balanced currents and voltages within the limits of inductive exposures) will be analyzed.

The balanced voltages in a three-phase power system form a symmetrical set of vectors equal in magnitude and 120 degrees apart in phase or may be readily analyzed into two such symmetrical sets of vectors. In either case, of course, the vector sum is equal to zero. In spite of this symmetry of voltages the induction to another conductor from the three balanced voltages is not necessarily zero since the coupling between each power wire and any other wire such, for example, as a wire of a telephone circuit, depends largely on its position with respect to such other wire. Since the spacings of the power conductors must be sufficient to provide adequate insulation, the distances from the various power conductors to the telephone conductor will usually be different and the inductions from these conductors will, therefore, be different and will not total zero. If the positions of the power conductors are rotated 120 electrical degrees periodically, however, the induction from the balanced components tends to be neutralized in each three successive equal lengths since the telephone line is thus exposed equally to all of the power wires. Such an arrangement of three successive equal lengths with two transpositions between them is called a transposition "barrel." The action of a barrel in neutralizing induction into adjacent circuits due to balanced voltages is illustrated in Fig. 5. It can be seen from this figure that the phase of the induc-

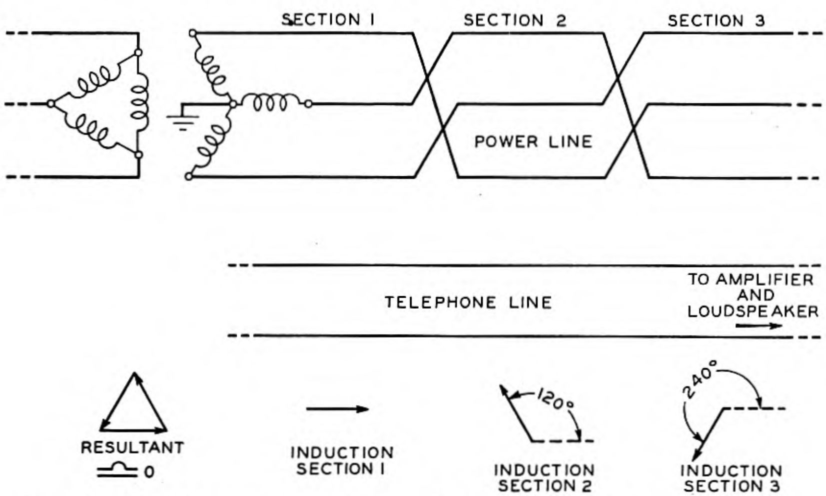


Fig. 5—Effect of power transpositions on induction due to balanced voltages.

tion into an adjacent circuit is rotated 120 degrees by each transposition so that in three sections the vector sum of the inductions would become zero if the inductions from the sections were identical in magnitude and exactly 120 degrees apart in phase. As a general rule, however, the actual inductions from the different sections are not identical in magnitude nor exactly 120 degrees apart in phase because of irregularities in the pole spacing and dimensions of the parallel and because of the fact that electrical waves take finite times to be propagated over the wires and hence do not have the same phase in successive lengths. For the usual distances encountered, the phase shift at fundamental frequency is small but it may be appreciable for the higher harmonic frequencies.

The analysis outlined above for balanced voltages can also be employed for balanced currents. When the load on the power line is not symmetrical the balanced currents will not be equal in magnitude and exactly 120 degrees apart in phase even though the vector sum is zero. However, these line currents may readily be divided into two sets of currents each of which may be represented by a set of vectors of equal magnitude and 120 degree phase displacement. The induction from each set of vectors may be neutralized by power transpositions (subject to the same limitations as for balanced voltages) and it follows, therefore, that the induction will be neutralized for their combination.

Transpositions in power systems affect the induction from residuals only to such extent as they may affect the magnitude of the residual voltages and currents (by providing better balance to earth). This is because the residuals act on the wires in parallel (as pointed out previously) so that interchanging the positions of the wires will not directly affect the inductive field about them.

To demonstrate the effect of power circuit transpositions on induction due to balanced and residual voltages, the miniature power circuit can be transposed to form a complete barrel. When the power circuit is energized with balanced voltages, a substantial reduction in noise from the loud speaker occurs when the transpositions are cut in. When the line is energized with residual voltage, however, cutting in power circuit transpositions does not cause any change in the noise from the loud speaker. In actual exposures, both balanced and residual voltages and currents may be present so that the effectiveness of power circuit transpositions will depend upon the particular conditions in each specific case.

Transpositions in Telephone Circuits

As in the case of power circuits, telephone transpositions have, from the standpoint of noise, two functions. The first is the equalization of admittance unbalances to earth and to other conductors, of the conductors of the particular circuit under consideration. The second is the reduction of noise due to direct metallic-circuit induction. (A third purpose, which is closely allied with the first and second, is the limitation of crosstalk coupling between the various telephone circuits on the same line.)

Within an inductive exposure, slightly different voltages may be induced on or along the two wires of a telephone circuit as pointed out previously. By transposing the wires frequently, they can both be exposed to the power system more or less equally and the voltages induced in them will tend to be equalized. The difference and hence the noise-metallic due to direct metallic-circuit induction thus is reduced. This is illustrated in Fig. 6. If the induction on the two sides of a transposition is identical in magnitude and phase, complete neutralization can be secured. In actual cases, however, these voltages are not identical in magnitude and phase because of irregularities in the exposure, irregularities in pole spacing, etc., and because of the phase shift and attenuation which were discussed in connection with power system transpositions.

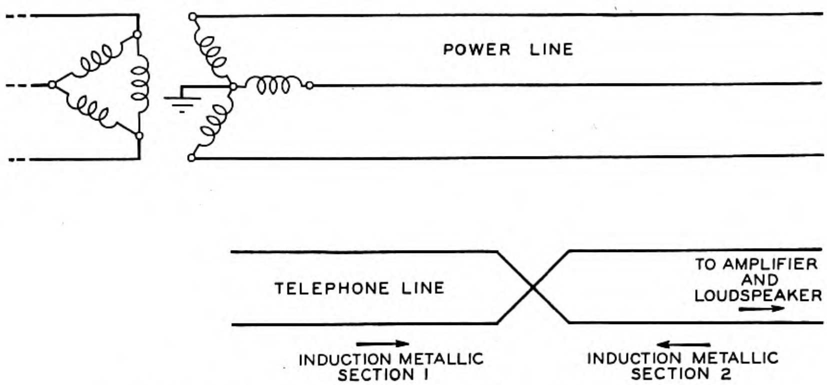


Fig. 6—Effect of telephone transposition on metallic noise.

Since the voltage to ground and the longitudinal circuit current due to either electric or magnetic induction, act on the telephone wires in parallel, telephone transpositions do not reduce them.

To demonstrate the effects of telephone circuit transpositions, the miniature telephone circuit is transposed. It is noted from the decrease in the noise from the loud speaker that, when the telephone circuit does not contain high resistance joints or other important unbalances, a substantial reduction in the noise metallic occurs when the telephone transpositions are cut in. However, no effect can be noted on the noise to ground.

Coordination of Transpositions

In order to summarize the effects of power and telephone circuit transpositions, Fig. 7 has been prepared. While this table applies only to transpositions within an exposure, it will be recalled that telephone and power system transpositions outside of exposures may have an important bearing on the balance of the circuits.

Transpositions	Induction From	Effect on Telephone Noise Met	To Ground
Power	Balance V.	Yes *	Yes
Power	Residual V.	No	No
Power	Balance I.	Yes *	Yes
Power	Residual I.	No	No
Telephone	All Types	Yes	No

* Power transpositions will reduce metallic noise on untransposed telephone lines. With telephone lines transposed the effects of power transpositions on metallic noise due to direct induction may be small.

Fig. 7—Summary of effects of transpositions within inductive exposures.

In some cases, it may be desirable to reduce not only the noise-metallic due to direct metallic-circuit induction but also the longitudinal-circuit noise due to balanced currents or voltages. An inspection of the table indicates that this may be done by transposing both the power and telephone circuits. In order to secure the greatest value from the transpositions in such cases they should be installed in such a way as to effectively "coordinate" with each other. In such coordinated layouts, the power circuit transpositions (where used) are largely relied on for reducing the longitudinal-circuit noise on the telephone circuits due to induction from balanced components and the telephone transpositions are largely relied on for minimizing the noise-metallic due to direct induction between the wires. Fig. 8 is a schematic diagram illustrating the principle of coordinated transpositions. It will be noted that the following considerations have been adhered to:

1. The telephone circuits are balanced, that is, both wires occupy both pin positions for equal lengths, between successive power circuit transpositions. This is necessary in order to ensure as close an approach as practicable to equality of induction on both sides of each telephone transposition.
2. The power circuit is transposed in a complete barrel. If the exposure is long or irregular, more than one barrel might be required.

In multi-wire telephone lines, the telephone transpositions are, of course, much more complex than those illustrated in Fig. 8, but in the systems designed for use in inductive exposures, so-called "neutral" points are established between which the circuits may be subjected to a uniform exposure. Consequently in a coordinated system of transpositions, it is ordinarily desirable that the neutral points in the telephone transposition system fall opposite or nearly opposite transpositions in the power system or other important electrical changes in the power system or in the exposure.

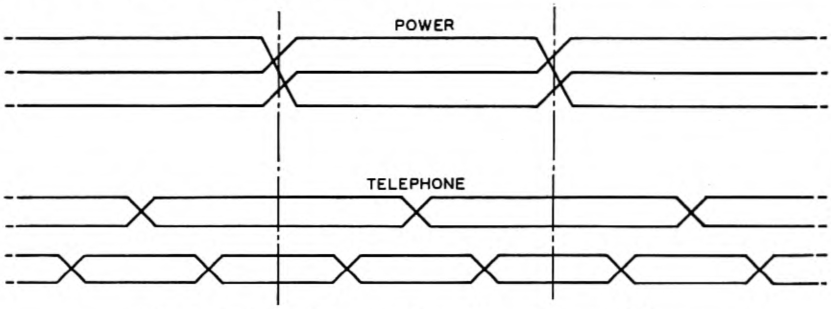


Fig. 8—Schematic layout of coordinated transpositions.

To illustrate the above, the demonstration apparatus is arranged to secure a coordinated layout. When the coordinated layout is cut in, only a relatively small amount of noise from the loud speaker is heard, and it is observed that the insertion of small series or shunt unbalances in the telephone circuit does not materially increase this noise (i.e., the telephone circuit is not particularly critical as regards unbalances) as long as the supply system is energized by balanced voltages only. When residual voltage is used on the miniature supply line, the longitudinal-circuit noise on the telephone system is higher and the telephone circuit is more critical as regards unbalances.

Inductive Influence of Power Lines

In considering some of the factors affecting the inductive influence of power lines, it should be recalled that, theoretically, a power system could be so constructed that it could set up no external electric or magnetic fields and consequently would have negligible influence. It is, as previously mentioned, impracticable to construct power lines in this way and consequently, the factors controlling the deviations from this condition require consideration.

Among the factors affecting the inductive influence of a power line are the amount of line current, the operating potential, the configuration of the wires, etc. It does not seem necessary to demonstrate these, but there are two additional factors of importance, as follows, which will be discussed:

1. The wave shape of the currents and voltages.
2. The magnitude (and wave shape) of residual voltages and currents. (Residuals were discussed briefly in connection with inductive coupling.)

Wave Shape

It is recognized as commercially impossible to build rotating machinery entirely free from harmonics. It is further recognized that some distortion of wave form is inherent with power transformers which must employ iron in their magnetic circuits. Harmonics are of interest from the standpoint of noise induction, since they may induce voltages of frequencies within the range ordinarily used in telephone message circuits. Induced voltages at such frequencies have much greater interfering effects (from the standpoint of noise) than does the voltage normally induced at the fundamental frequency. The approximate relative interfering effects of voltages of different frequencies in typical telephone circuits are shown in Fig. 9 which is a so-called "noise weighting" curve.

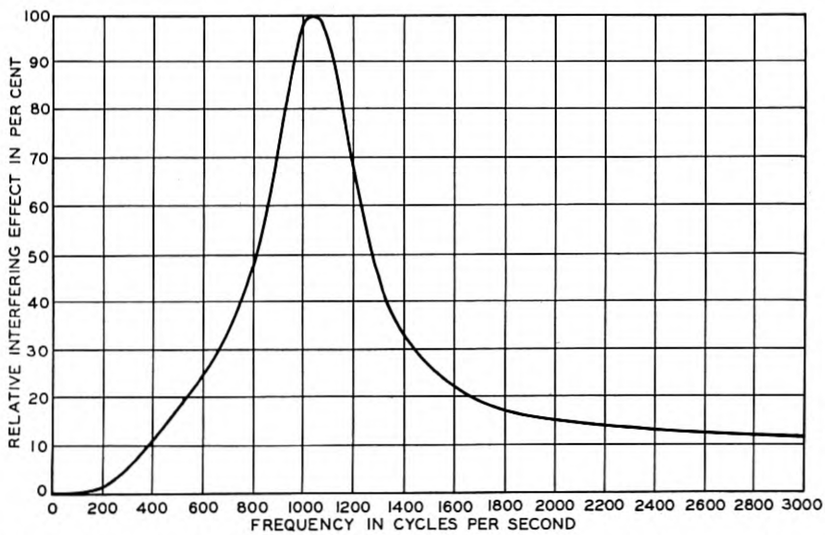


Fig. 9—Curve showing approximate relative interfering effects of voltages of different frequencies across a telephone circuit.

The demonstration set-up for impressing voltages of two different wave shapes on the untransposed power line is shown in Fig. 10. With the switch in the "normal" position, the wave shape is that taken directly from the commercial power supply. A wave shape of voltage having greater harmonic content than that of the commercial voltage, can be secured by throwing the switch to "distorted." The operation of the circuit is then as follows:

1. The commercial power supply is connected to the 10-volt windings of the transformers through balanced resistances which are so proportioned that the voltage drop due to the magnetizing current is sufficient to reduce the voltages across the windings to about 10 volts.
2. The resistances form such a large proportion of the total impedances presented to the incoming circuit that the currents through the windings are controlled almost entirely by them and, since they are non-inductive, this current has approximately the same wave shape as the voltage of the power supply. Therefore, since the magnetizing harmonics cannot appear to any large extent in the magnetizing current, they appear in the voltage across the transformers and the voltage wave is, therefore, distorted.
3. The distorted voltage wave on each transformer is stepped up between the 10 and 115 volt windings and is impressed on the line at about 115 volts to neutral.

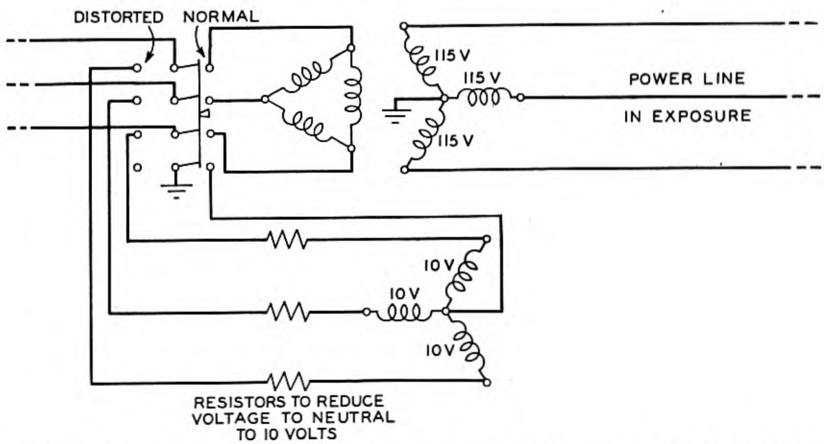


Fig. 10—Arrangement for comparing the inductive influence of balanced voltages of different wave shapes.

Figure 11 is an oscillogram showing the "normal" and "distorted" wave forms and it will be noted that they have about the same r.m.s. values although the distorted wave is much more irregular indicating the greater harmonic content. When the switch is thrown from "normal" to "distorted," the noise from the loud speaker increases and its characteristic sound is changed, indicating the effects of increasing the harmonic content of the voltage wave.

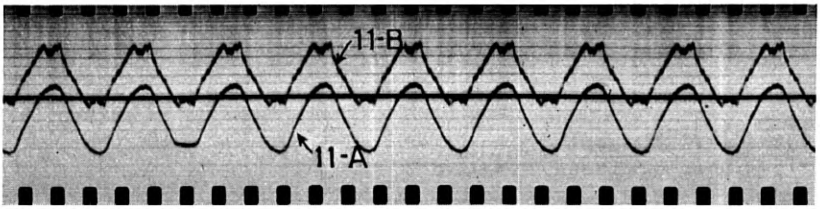


Fig. 11—Oscillograms of normal and distorted balanced voltages.

11-*A*—Impressed voltage,
11-*B*—Balanced distorted voltage to neutral.

In practice, harmonic voltages and currents may arise not only from generating and transforming equipment but also may occasionally arise from some particular load equipment such, for example, as certain types of rectifiers or rotating machinery.

Residual Voltages and Currents

The inductive influence of a voltage or current of a given magnitude and wave shape depends to a considerable extent on the dimensions of the circuit in which it acts. For balanced currents or voltages (or balanced components of the actual currents or voltages on line wires), which, as discussed before, are confined to the wires of the power circuit, the dimensions of the circuit are much smaller than for the residual currents or voltages which involve the earth as part of their circuit.

In order to illustrate the relative inductive influences of a given magnitude and wave shape of voltage, when acting in a balanced manner and as a residual, the miniature power line is energized in two different manners. First (the normal manner) the voltage is impressed on it through a bank of transformers connected "delta" on the supply side and "Y-grounded" on the line side. With these connections, the voltages impressed on the three line wires are approximately equal and 120 electrical degrees apart and thus are closely balanced. Next, using the same transformer connections, the line wires are energized in parallel to earth and consequently, the vector sum (residual) is equal to three times the normal phase-to-neutral voltage. The power circuit connections used are shown in Fig. 12 and the telephone circuit connections used are the same as shown in Fig. 4-*A*. The increase in the noise from the loud speaker when residual voltage is used shows that the influence of the power line is greater under these conditions.

In addition to the effect of residuals in increasing the inductive influence of a power line, the induction due to residuals is not affected by transposing the power line (as was pointed out in connection with the discussion of coupling).

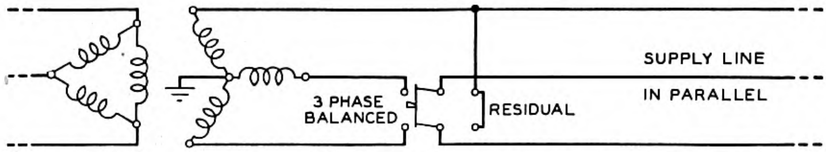


Fig. 12—Arrangement for comparing balanced and residual voltages or currents.

It may be of interest to examine briefly some of the causes of residual voltages and currents. For example, in a three-phase system, harmonic currents or voltages-to-neutral which are odd multiples of three times the fundamental frequency are in phase in all three line wires and hence tend to be residual. Such triples can be present in appreciable amounts only with certain types of power apparatus and connections.

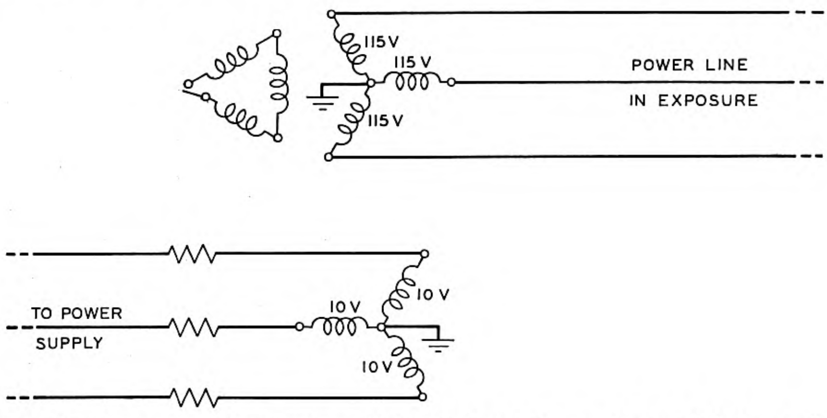


Fig. 13—Arrangement for showing added inductive influence due to triple harmonic voltages.

Perhaps the most important condition giving rise to triple harmonic frequency residual currents or voltages is the connection of grounded-neutral Y-connected generators which have triple harmonic voltages between line and neutral, directly or through Y-Y connected transformer or Y-connected auto-transformer banks (with no or small tertiary windings, and with grounded neutrals) to power lines. The use of Y-Y banks may also cause triple frequency residuals on the lines due to the magnetization characteristics of the transformers themselves although when used with Y-connected grounded generators, the transformer effects are usually less important than the generator effects (unless the "triples" in the generator are unimportant or are suppressed).

To demonstrate the effect of triple harmonic currents, the arrangements shown in Fig. 13 have been set up. This set-up is similar to that used in showing the effect of differences in wave shapes of balanced voltages except that, to show the added effect of triple harmonic residuals, the delta winding is opened. This removes the path for triples to circulate within the transformer bank and permits them to be impressed on the line. Figures 14-*A* and *B* are oscillograms showing the effect on the wave shape of the voltage to neutral of opening the delta. The noise from the loud speaker increases when the delta is opened showing that the triple harmonics cause an increase in the influence of the power line.

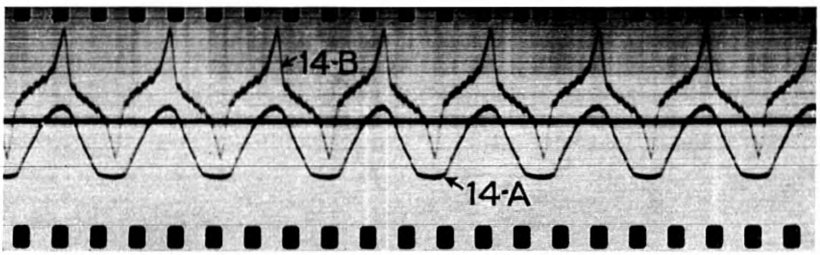


Fig. 14—Oscillograms showing voltage wave shape including triple harmonics.
14-*A*—Impressed voltage,
14-*B*—Distorted voltage to neutral, including triple harmonics.

An interesting demonstration showing the relation as regards residuals of triple and non-triple harmonics on an otherwise well balanced three-phase system can be performed as follows:

1. The untransposed power line is first energized with balanced distorted voltages as described previously. The amount and character of the noise are observed closely.
2. Triple harmonic voltages are added by opening the delta winding on the transformer bank. Under these conditions, the induction from both the triple and non-triple harmonics can be recognized by the differences in the character of the sounds.
3. The power line is now transposed and the noise due to the non-triples practically disappears leaving the noise from the triples unaffected.

This illustrates the residual character of the triples since, as shown previously, the power system transpositions do not affect the induction from residuals.

Single-Phase Extensions

One of the special conditions under which residual currents or voltages (particularly of the non-triple series of harmonics) are set up on a power system is where single-phase circuits are connected metallically to 3-phase circuits. With such a connection, the inductive influence of both the single-phase and 3-phase parts of the power circuit may be affected. Briefly the conditions are as follows:

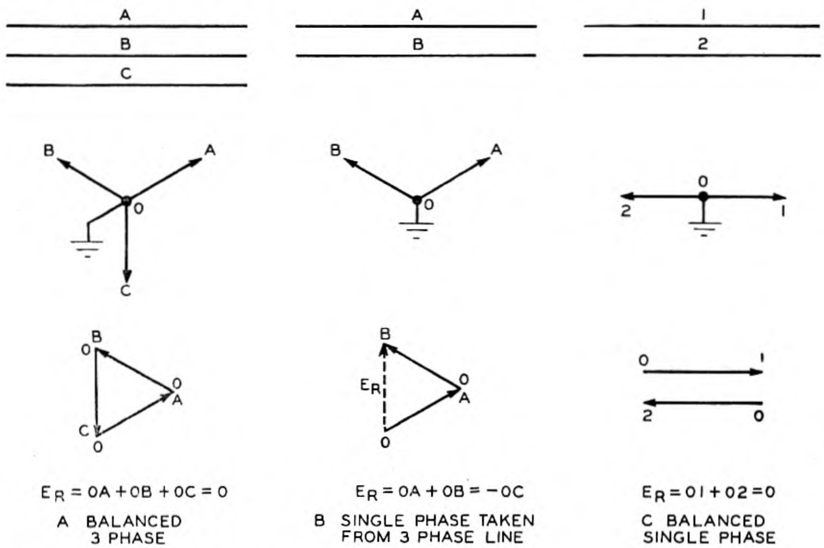


Fig. 15—Comparison of residual voltages in perfectly balanced 3-phase line; a single-phase tap from 3-phase line, and a perfectly balanced single-phase line.

Single-phase portion

1. On the single-phase portion of the circuit, a residual voltage exists which ordinarily is approximately equal to the normal voltage to ground of a phase wire. This is readily evident from an inspection of the vector relations shown on Fig. 15-*B*. Fig. 15-*C* shows that there is nothing inherently unbalanced in single-phase circuits; it is only when they are connected directly to a three-phase circuit or have some unbalanced connections that they have residuals on them.
2. Figure 16 shows schematically the arrangements used to illustrate the effects of metallically connecting a single-phase circuit to a three-phase circuit. By throwing the four-pole, double-throw switch, the noise to ground in the miniature telephone circuit (exposed only to the single-phase circuit) with the single-phase

portion isolated from the three-phase portion by a transformer and with it metallically connected can be compared. With the transformer connected (thereby creating a condition similar to Fig. 15-*C*) the noise in the loud speaker is much lower than when a metallic connection is used and thus indicates a substantial reduction in the residuals.

3. The demonstration setup is so arranged that the single-phase portion can be transposed. With the single-phase portion metallically connected to the three-phase portion, transposing the single-phase portion causes relatively little change in the noise from the loud speaker. However, when the single-phase portion is isolated from the three-phase portion by the transformer, transposing it further reduces the noise materially. When the single-phase portion is connected metallically to the three-phase portion, the induction is largely due to residual voltage and as such is not affected by the power circuit transpositions. When it is connected through the isolating transformer, however, there is no residual voltage present and the induction, being due to balanced voltages, is materially reduced by the power transpositions.

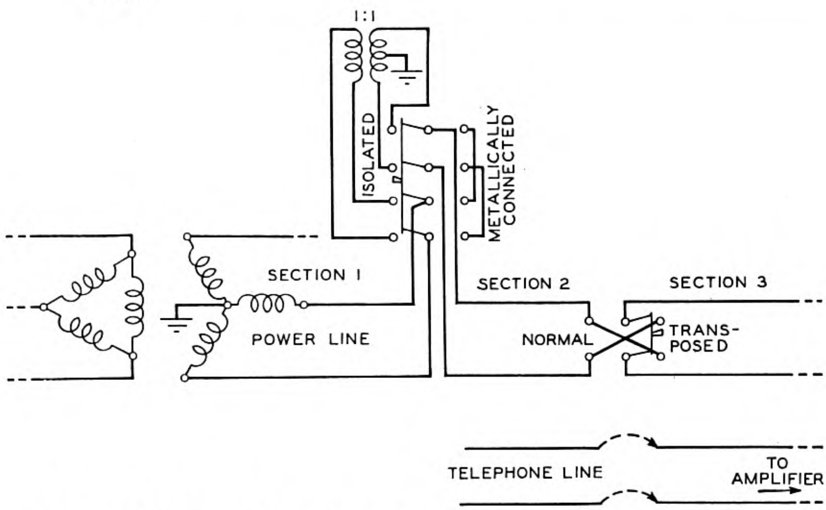


Fig. 16—Influence of single-phase extension to three-phase power line.

Three-phase portion

1. As far as the three-phase portion of the line is concerned, the single-phase extension acts as additional admittance to ground on two of the wires. Consequently if the single-phase extension is long,

the admittance unbalances between the various wires and ground may be fairly large.

2. In considering the effects of the admittance unbalances, there are two conditions which must be considered; where the transformers supplying the three-phase portion are "Y grounded" on the line side, and where they are "delta" on the line side. When the supply transformers are connected delta on the line side, there is no path for residual current into the transformers and the voltages of the conductors to earth adjust themselves so that the net charging current to earth is zero (although there will be some interchange of charging current between various portions of the network). This condition requires unequal voltages to earth, the voltages of the wires having the higher capacitances being lower than those of the lower capacitance wires. This generally gives a residual voltage.
3. When the supply transformers are connected Y-grounded on the line side, the voltages of the wires to ground are controlled by the transformer voltages and the principal effect of a single-phase extension is a tendency to cause residual current.

The discussions above apply particularly to power systems which are electrically short at all of the important harmonic frequencies present. If the systems are long enough so that propagation effects (particularly "quarter wave-length" effects) must be considered at any of the important harmonic frequencies present in the voltage or current waves, these simple analyses must be modified. These propagation effects cannot be demonstrated with the apparatus available and will not be discussed further except to point out that they are not infrequently encountered in field problems.

Inductive Susceptiveness of Telephone Circuits

As pointed out previously, theoretically a telephone circuit could be constructed so that it would not be affected by any fields which would be set up by nearby electrical systems and hence would have zero susceptiveness. However, as in the case of the power line, it is not practicable to build such ideal telephone lines and consequently, the consideration of telephone lines in inductive exposures has to do with the deviations from perfection in this respect.

As was indicated earlier in this article, the metallic type of telephone circuit is now usually used. The grounded system which uses one wire with earth return, was employed exclusively in the very early days and is still used in some cases, particularly in sparsely settled areas.

The grounded circuit represents completely unbalanced conditions since the sides of such a circuit have a separation comparatively great compared to that of a metallic circuit. Consequently, the inductive susceptiveness of a grounded circuit is much greater than that of a metallic circuit, even if the latter is not transposed. Furthermore, a grounded circuit cannot be transposed practicably. To illustrate the difference in the susceptiveness of the two types of circuits, the telephone circuit of the demonstration set up has been arranged as shown schematically in Fig. 17 so that either of the two types of circuits may be obtained. The power circuit arrangements are as shown in Fig. 4-*B*. The large reduction in the noise from the loud speaker which occurs when the connections are changed from grounded to metallic, shows the decreased susceptiveness of the latter type of circuit.

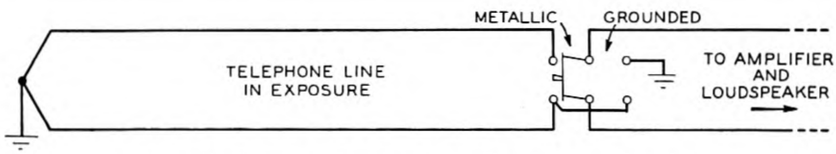


Fig. 17—Comparison of noise in metallic and grounded circuit.

For metallic circuits, the inductive susceptiveness depends on a number of factors such, for example, as the spacing of the wires, the power levels, and the circuit balance. Some of these are discussed below.

Spacing

Since the direct metallic induction (which, as discussed before, is a function of the difference of the voltages induced on or along the two sides of the circuit) is about proportional to the distance between the two sides of the circuit, this separation is of interest from the standpoint of the circuit susceptiveness. The smaller the spacing of the wires, all other things remaining the same, the smaller ordinarily will be the direct metallic induction and the noise-metallic from this source.

Power Level

Another important element in determining the inductive susceptiveness of a telephone circuit is the power level of the telephone waves transmitted over the circuit. The more powerful the telephonic currents at a point, the less they will be interfered with by a given amount of noise power which may be induced in the circuit at that point. This is particularly important on long toll circuits where the telephonic power level may be materially affected by the spacing, power carrying capacity and adjustments of the telephone repeaters usually used in such circuits.

Balance

In order that a telephone circuit may be perfectly balanced, the series impedances of the two sides must be identical in each element of length and the admittances of the two sides to earth and to other conductors likewise must be identical.

Since it is impracticable to construct telephone circuits of perfect symmetry, unbalances exist and these are classified as "series impedance" and "shunt admittance" unbalances. By a "series impedance" unbalance is meant a difference between the series impedances of the two wires composing the circuit. Such an unbalance may be caused, for example, by a joint which does not have a negligible resistance. If a "bad" joint exists, the longitudinal currents due to the induced voltages encounter unequal impedances in the two wires. Consequently, the currents in the two wires tend to be unequal, the difference causing current through the terminal impedances and hence causing metallic circuit noise. The effect of a high resistance joint depends upon the magnitude of longitudinal current along the wires as well as the unbalance in resistance caused by the joint. To illustrate the effects of a high resistance joint, the demonstration set-up is arranged to minimize the noise-metallic due to direct induction (by transposing it) and the high resistance joint is then inserted. (See Fig. 18.) The large increase in the noise from the loud speaker indi-

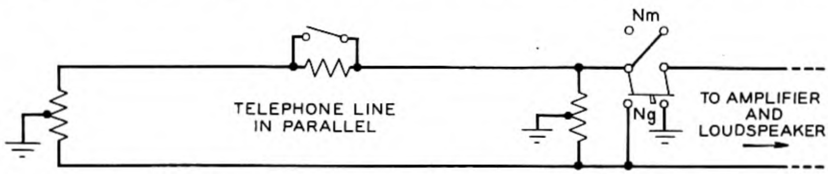


Fig. 18—Arrangement for showing effect of high resistance joint in telephone line.

cates the effect of the joint on the noise-metallic. On the other hand, listening to the noise-to-ground when the joint is inserted, one can detect no effect.

Admittance unbalances are generally due to either unbalanced capacitances or leakages to earth of the two wires. Such unbalances when acted on by the noise to ground cause more current to flow to ground from one side than from the other. Part of this current flows around the metallic circuit and causes noise-metallic. To illustrate the effect of an admittance unbalance, a small condenser or a high-resistance leak can be bridged between one wire of the telephone circuit and earth in the demonstration apparatus. As before, the effect of the unbalance on the noise to ground is negligible, but it may cause a material increase in the noise-metallic.

While a 2-wire metallic telephone circuit has been used in the discussions, the same principles apply to a phantom circuit. In considering the effects of unbalances, transpositions, etc., on phantom circuits, the two wires composing each of the side circuits from which the phantom is derived may be considered as being in parallel and treated as if they were single conductors. With this method of treatment, the discussions of a 2-wire circuit can also be applied to a phantom circuit, bearing in mind, among other things, that with four wires to treat with instead of two, an unbalance in any of the four wires will react on the phantom circuit as well as on the side circuit of which it is a part.

While for simplification the demonstration has been confined to the effects of unbalances in the line conductors, it is evident that similar effects can result from the equivalent series or shunt unbalances in terminal equipment in central offices, in subscribers' sets, cables, etc.

Interconnection of Balanced and Unbalanced Telephone Circuits

One of the factors which is of interest in connection with noise on telephone circuits is that which is concerned with the phenomena which occur when a well balanced and a poorly balanced telephone circuit are connected together. It was pointed out previously that a well balanced and transposed telephone circuit may be relatively quiet even if it is exposed to induction. Also, if a poorly balanced circuit is not exposed to induction, it may be quiet. If, however, the exposed, well balanced circuit and the unexposed, poorly balanced circuit are connected together either at some point along the line or through a cord circuit not containing an isolating repeating coil, the overall connection may be noisy since the interconnection in effect unbalances the otherwise well balanced circuit.

To demonstrate this the conditions shown in Fig. 19 are set up. The metallic portion of the circuit at the left of the diagram is exposed to the 3-phase power line but is well transposed and balanced. The grounded circuit, shown at the right of the diagram, is not noticeably exposed.

The noise heard when the loud speaker is connected to the metallic circuit (although it is exposed) is relatively low. Likewise, the noise on the grounded circuit is relatively low. When, however, the grounded circuit is connected to the metallic circuit the noise on the overall circuit immediately rises because of the unbalancing effect of the grounded circuit.

It will be recognized that the general principles involved in this last demonstration are essentially the same as those which were involved in the demonstration of the effect of a single-phase extension to a 3-phase

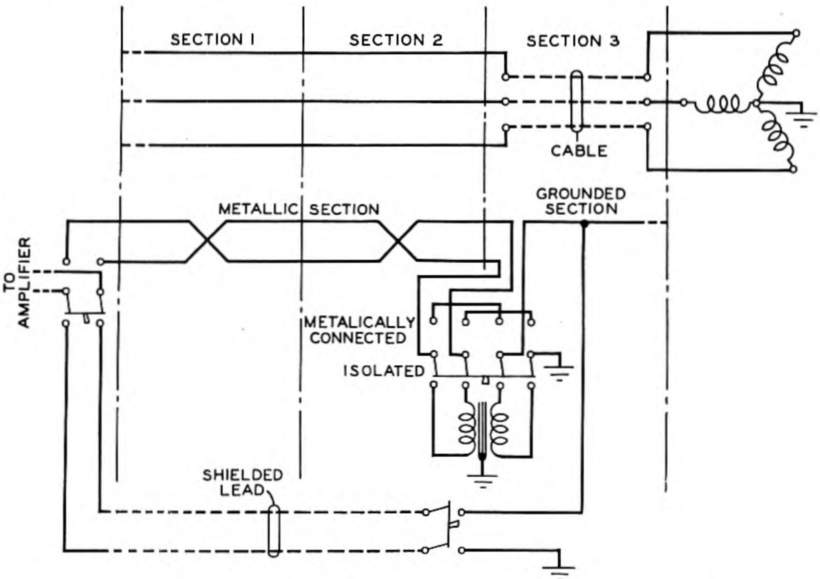


Fig. 19—Effects of interconnecting metallic and grounded circuit.

power line. In the case of the single-phase extension, it was possible to reduce the inductive influence by isolating the single-phase part from the 3-phase part by means of an isolating transformer. Following the same line of reasoning, it should be possible to reduce the effect of the connection between the metallic and grounded parts of the telephone circuit by means of an isolating transformer. Inserting a repeating coil between the metallic and grounded portions provides such isolation and it is noted from the reduction in noise when this repeating coil is inserted, that the conditions are essentially the same as when the grounded portion is disconnected from the metallic portion. (This whole analysis and demonstration, of course, applies only when the grounded portion is unexposed since the grounded circuit is totally unbalanced and hence would quite likely be noisy if it were subjected to direct induction.)

Carrying the similarity of these two demonstrations a step farther, it will be recalled that it was shown that when the single-phase and 3-phase portions of the power circuit were metallically connected, transposing the single-phase portion resulted in relatively small reduction in the inductive influence because the induction was primarily a function of the residuals on the line. Similarly when the metallic and the grounded circuit are metallically connected, it is observed that the transposing of the metallic circuit produces a relatively small reduction in noise. However, if the repeating coil is inserted between the metallic and grounded circuits it is observed that transposing the metallic portion materially reduces the noise on the overall connection, since the transpositions reduce direct induction in the metallic circuit and the noise to ground is not given an opportunity to react on the unbalances.

Audio Frequency Atmospherics *

By E. T. BURTON and E. M. BOARDMAN

Various types of musical and non-musical atmospherics occurring within the frequency range lying between 150 and 4000 c.p.s. have been studied. Particular attention is directed to two types of the former, one a short damped oscillation, apparently a multiple reflection phenomenon, and the other a varying tone of comparatively long duration, probably related to magnetic disturbances. Several quasimusical atmospherics which appear to be associated with the two more distinct types are described. Dependence of atmospheric variations on diurnal, seasonal and meteorological effects is discussed. Characteristics of audio frequency atmospherics are shown in oscillograms and graphs.

Introduction

IN connection with a study of communication problems, observations of submarine cable interference were made over periods totaling about 20 months during the years 1928 to 1931. These experiments were conducted at Trinity, Newfoundland; Hearts Content, Newfoundland; Key West, Florida; Havana, Cuba; and at Frenchport, near Erris Head, Irish Free State. A few supplemental measurements of audio frequency atmospherics received on large loop antennas were made in 1929, 1931 and 1932. These experiments were made at Conway, New Hampshire, at two locations in New Jersey and in Newfoundland. Work carried out at the Newfoundland and New Hampshire locations has been commented upon in previous reports.[1]

Since, for the most part, industrial and communication interferences were of small magnitude at all locations, it has been possible to select for presentation data confined to atmospherics. These data will be limited mainly to the frequencies between 150 and 4000 c.p.s., although measurements were made over the range from 40 to 30,000 c.p.s.

The principal apparatus used at each location consisted of an especially designed vacuum tube amplifier with which all other apparatus was associated. The overall gains of the amplifiers used at the various locations varied somewhat according to the conditions to be met, the frequency characteristics being adjusted approximately complementary to that of the pick-up conductors. The Ireland amplifier consisted of seven transformer coupled stages grouped to form three units. The impedance at the junction points of units was

* Presented at U. R. S. I. convention, Washington, D. C., April 27, 1933. *Proc. I. R. E.*, 21, p. 1476, October, 1933.

[1] E. T. Burton, "Submarine Cable Interference," *Nature*, 126, p. 55, July 12, 1930; and E. T. Burton and E. M. Boardman, "Effects of Solar Eclipse on Audio Frequency Atmospherics," *Nature*, *131*, p. 81, January 21, 1933.

600 ohms to facilitate insertion of attenuators and filters. The maximum gain for the three amplifier units was 200 db, attenuators and filters being used at all times to control the output intensity. The amplifier was designed to minimize noise, inherent in such apparatus, and to be highly stable throughout long periods of practically continuous operation.

In addition to several high-pass and low-pass filters, 17 narrow band filters designed to cover in small steps the range from 150 to 3800 c.p.s. were available. A filter switching panel was used to facilitate observations of various frequency ranges in rapid succession.

The output was arranged to supply various recording and indicating devices. R.m.s. measurements were made by means of a thermocouple with a long period direct reading and recording meter. A device employing three-element gas-filled tubes was used to measure peak voltages. A magnetic recorder was employed in securing a few sound records of atmospherics. Oscillograms which are shown in this article were subsequently prepared from these records. The Ireland amplifier with some of its associated apparatus is shown in Fig. 1.

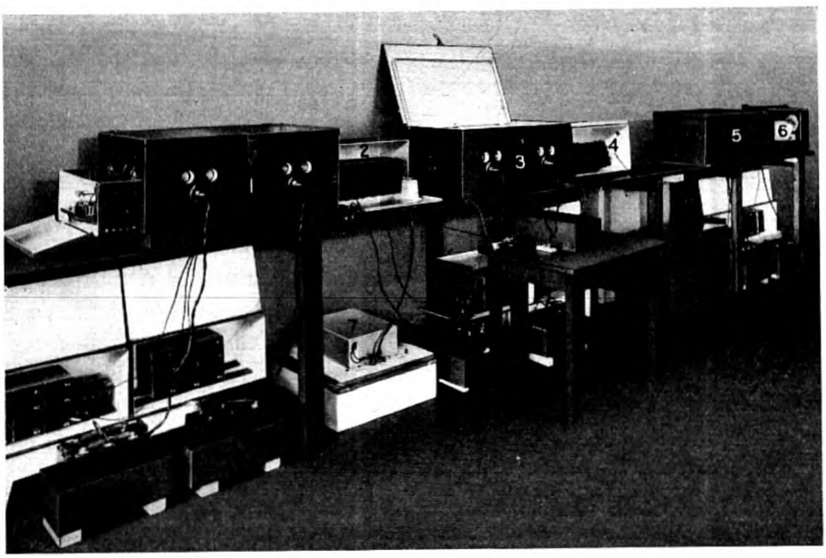


Fig. 1—Amplifier and associated apparatus used at Frenchport, Ireland.
(1) First amplifier unit
(2) 1st Attenuator
(3) 2nd Amplifier unit
(4) 2nd attenuator
(5) 3rd amplifier unit
(6) Recorder
(7) Band pass filter

The amplifier with each of the filters taken separately was calibrated with input supplied by the thermal agitation in standard resistances ranging from 50 to 250 ohms. The calibration temperature was approximately 23° C. Check calibrations were made weekly and at such times as changes were made in the apparatus. The stability of the entire system was such that over periods of months measurements were made with an accuracy closer than $\pm$ 1/2 decibel.

In interpreting data on atmospherics of low amplitude, such as received on submarine cables, it is necessary to take into account the random voltages generated in the amplifier circuits and the thermal agitation voltages of the conductor connected to the amplifier input. Both of these voltages appear in the output circuits mingled with the amplified atmospherics. The former originate principally in the first stage of the vacuum tube amplifier. Thermal agitation produces a random voltage, uniformly effective at all frequencies. The r.m.s. amplitude of this voltage is dependent upon the frequency range considered, the resistive component of the impedance of the conductor and the temperature of the conductor.[2] The conductor in this case is the cable or antenna circuit. The r.m.s. values of these voltages, when integrated over periods of time comparable to those occupied in taking data on atmospherics, are substantially steady; therefore, their separation from the atmospheric voltages is not difficult. Corrections for both amplifier and thermal noises have been made on the data presented.

Observations of audio frequency atmospherics received on long antennas and loop aerials have been reported by several observers.[3] Their accounts describe the general characteristics, although some confusion has occurred in identification of the musical atmospherics. In view of the fact that the apparatus used by us was particularly adapted to reception and analysis of frequencies in the audio range, it appears that our data may add considerably to the information previously disclosed.

Types of Atmospherics

Audio-frequency atmospherics observed on submarine cables are essentially the same as those received from a long antenna except for high attenuation and frequency discrimination attributable to the cable characteristics and to the shielding effect of sea water.[4] The low

[2] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, *32*, p. 97, July, 1928.

[3] H. Barkhausen, "Whistling Tones from the Earth," *Phys. Zeits.*, 20, p. 401, 1919. T. L. Eckersley, "Electrical Constitution of the Upper Atmosphere," *Nature*, *117*, p. 821, June 12, 1926.

[4] John R. Carson and J. J. Gilbert, "Transmission Characteristics of Submarine Cables," *Jour. Franklin Inst.*, *192*, p. 705, December, 1921.

frequencies, when observed on a submarine cable, are of comparatively high amplitude, appearing as a deep rumble intermittently broken by noises variously described as splashes and surges. The range from 500 to 1500 c.p.s. generally consists largely of clicks and crackling sounds which accompany the low-frequency surges. At times substantial amplitude increases occur accompanying quasi-musical sounds, which may dominate this frequency range. In the upper voice range intermittent hissing or frying sounds are observed, often accompanying surges in the low-frequency range. Above 1800 c.p.s. occur at least two ranges which at times possess slight tonal characters. In addition to the slightly musical sounds, two varieties of distinct musical atmospherics have been observed and given the onomatopœic names "swish" and "tweek." Particular interest attaches to these because of their extraordinary character.

Diurnal and Seasonal Characteristics

The daytime non-musical atmospherics consist ordinarily of intermittent low-amplitude impulses. As a general rule the night-time intensities are considerably higher; the impulses being more frequent and more prominent than during the daylight hours. The night intensity is further increased by the presence of the type of musical atmospheric known as tweek.

During a usual day, the intensity of audio-frequency atmospherics from sunrise until mid-afternoon is comparatively low. During the afternoon, a slow rise may or may not occur. Shortly following sunset, a gradual increase of intensity is usual. This rise continues for two hours or more after which a high level is maintained rather consistently until shortly before daybreak. A brief increase sometimes occurs at this time followed by a steady decrease, the daily minimum being reached usually shortly after sunrise.

Fig. 2 shows examples of summer and winter audio-frequency atmospheric intensities over 24-hour periods. While these curves show the usual characteristics, extraordinary conditions may result in wide variations. The occurrence of local electrical storms or intense disturbances of the earth's magnetic field usually contribute markedly to these anomalies.

The diurnal amplitude variations of certain types of atmospherics may be reasonably explained by assuming the continued presence of an audio-frequency reflecting layer in the upper atmosphere, and assuming a low lying ionized attenuating region [5] to be present during

[5] Such a region affecting radio frequencies is described by R. A. Heising, *Proc. I. R. E.*, **16**, p. 75, January, 1928.

daytime only. During the sunlight hours, disturbances occurring in the vicinity of the observation point may be received by direct transmission without unusual attenuation. Atmospherics of distant or high origin should suffer considerable attenuation in passing through

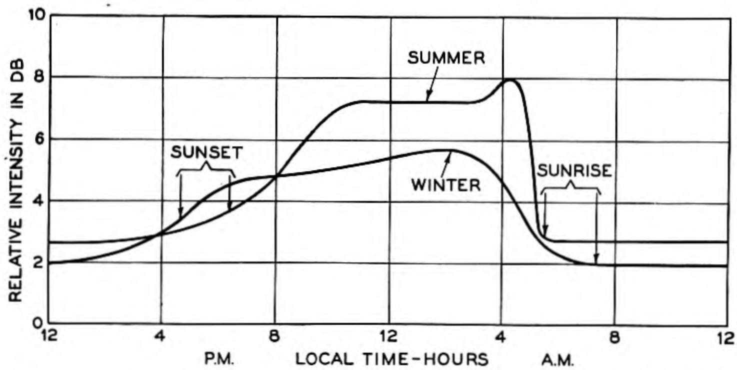


Fig. 2—Typical diurnal intensity curves, for frequency range from 150 to 3000 c.p.s.

the damping region. Following sunset, the damping ionization may be expected to gradually dissipate, resulting in a slow increase of the static intensity as transmission from the upper atmosphere and from horizontally distant regions is improved.

It is probable that in the morning the damping ionization appears at a given point almost immediately upon arrival of the first direct sunlight, and that the transition period corresponds to the time required for the earth to rotate through an angle corresponding to that section of the damping region which may appreciably affect the atmospherics reaching the observation point.

Our observations have shown that the general intensity of the regularly occurring types of atmospherics increases in the spring, the rise beginning about March. During a period from possibly May to September, the intensity is comparatively high. During September and October a reduction occurs, and from the latter part of October until March the intensity is low. The periods as given above are approximate, since they are based on fractional year observations in all except one case.

Comparison of Fig. 2 with diurnal variation curves of Potter [6] for 50 kilocycles and 2 megacycles, and with seasonal variations presented by Espenschied, Anderson and Bailey [7] for 50 kilocycles shows definite similarities.

[6] R. K. Potter, "Frequency Distribution of Atmospheric Noise," *Proc. I. R. E., 20,* p. 1512, September, 1932.

[7] Espenschied, Anderson and Bailey, "Transatlantic Radio Telephone Transmission," *Proc. I. R. E., 14,* p. 7, February, 1926.

Tweeks

A tweek consists of a damped oscillation trailing a static impulse. Its audible duration appears to be less than 1/8 second and the initial peak amplitude may approximate that of the maximum audio frequency static impulses.

Oscillographic reproductions of sound records obtained in Ireland disclose that the tweeks practically always start above 2000 c.p.s. and reduce very rapidly toward a lower limiting frequency where a considerable portion of the time of existence is spent. In some cases the highest observed frequency at the beginning of a tweek was in the vicinity of 4000 c.p.s., which was the upper transmission limit of the apparatus. In Fig. 3 is shown an oscillogram of tweeks trailing

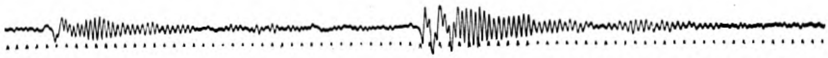


Fig. 3—Oscillogram of tweeks. Timing impulse frequency, 1000 c.p.s.

static surges. While in these tweeks, any initial high frequencies are obscured by the prominent static surge, some oscillograms have been made while using electrical filters to suppress the frequencies mainly responsible for the initial impulse. These oscillograms often showed

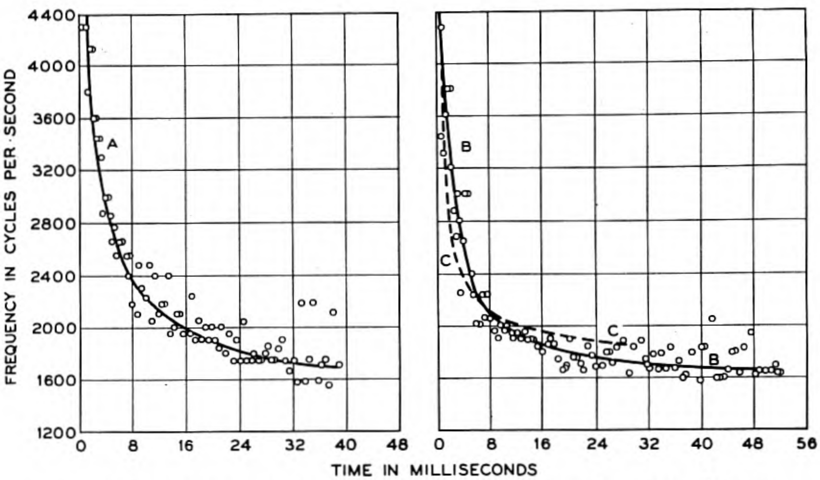


Fig. 4—A and B, tweek frequency variation curves. C, computed curve.

initial frequencies as high as 4000 c.p.s. Two tweek frequency determinations made from oscillograms are shown in Fig. 4. These illustrate the initial rapid frequency reduction and the subsequent gradual approach to a constant. While not an accurate definition, frequency, as determined from the oscillograms, is taken as the reciprocal of the time spacing of successive impulses. Due to the difficulty in accurately measuring these short time intervals, especially in the presence of other forms of atmospherics, there is a possibility of error which might account for the irregularities in the location of points. However, irregularities in effective height of the reflecting layer might be expected to produce a like result.

With one possible exception,[8] tweeks have never been observed by us during daytime except near sunrise and sunset. In the usual case, the intensity of static impulses increases during the early evening with no indication of tonal quality. At twilight certain of the impulses are observed to be accompanied by a slight indication of a highly damped frequency. Shortly thereafter the characteristic tweek tone appears, often trailing a good share of the static impulses. Both tweek rate and intensity ordinarily increase for some two hours. For the remaining hours of darkness the tweeks, usually of low damping, continue with many irregular variations in intensity. Just previous to the approach of daylight a brief increase in tweek rate often occurs followed by a rapid reduction in both intensity and rate of occurrence. The last highly damped tweek is usually observed several minutes before sunrise.

H. Barkhausen [9] in attempting to explain the type of atmospheric tone known as the "swish" or the "long whistler" considers the multiple reflection of an impulse. While our observations indicate this theory to fail in explanation of the swish, it appears to be applicable to tweeks. According to this theory a tweek may be produced by energy, from a source of momentary static disturbance, arriving at a receiving point as a series of impulses. The first impulse arrives by direct transmission. Shortly thereafter a second impulse arrives after having suffered one reflection at an ionized layer in the upper atmosphere. The third impulse arrives after two reflections from the ionized layer and one from the earth's surface. Other impulses follow in like manner. In case the origin of the disturbance is not near the observation point, the time spacing of the observed impulses results in a reducing frequency, initially varying rapidly and finally approaching an asymptotic value. The initial frequency is dependent upon the distance from source to observer and the reflecting layer height, while the lowest frequency depends upon the height alone. The failure of tweeks to appear in daytime may be attributed to damping by sunlight ionization at low altitudes. Occasional highly damped

[8] E. T. Burton and E. M. Boardman, "Effects of Solar Eclipse on Audio Frequency Atmospherics," *Nature, 131*, p. 81, January 21, 1933.

[9] H. Barkhausen, *Proc. I. R. E., 18*, p. 1155, July, 1930.

and weak tweeks observed before sunset or after sunrise probably originate at considerable distance respectively to the east or west within regions not exposed to sunlight.

The multiple reflection theory of tweeks, as explained above, concerns a single wave train originating in a disturbance located near one of the reflecting surfaces. It may be shown that an impulse originating anywhere in the intervening space might produce a similar effect, although the initial frequency would be altered by the location in altitude. Furthermore, were the point of origin well separated from both surfaces, two simultaneous wave trains differing somewhat in rate of frequency change would occur. Phasing effects, which might be attributed to this have been found in several oscillograms.

Based on the multiple reflection theory, the curve C in Fig. 4 was calculated assuming the point of origin to be located near the earth's surface. The altitude of the reflecting layer was taken as 83.5 km. (55 miles) and the distance between source and observer as 1770 km. (1100 miles). While this curve only roughly approximates the form of the tweek curves of Fig. 4, an explanation of the discrepancy may lie in a variation in effective layer height in accordance with the change in angle of incidence of the successive impulses. Such a relation in the case of radio frequencies has been described by Taylor and Hulburt.[10]

Comparison of the lower limiting frequencies of individual tweeks with an oscillator calibrated in small steps has shown at times an almost continual drift in frequency. This may be interpreted as a corresponding variation in the effective height of the reflecting layer. In one five-minute period during complete darkness, examination of 24 tweeks showed the lower limiting frequency to vary irregularly between 1690 and 1720 c.p.s. This indicates a variation in effective layer height between approximately 88.5 and 87 km. The variations of lower limiting tweek frequencies noted at our various observation points have indicated the reflecting layer to vary between 83.5 and 93.2 km. during the hours of complete darkness. No marked variations of mean tweek frequency, in respect to either season or latitude, have been observed.

During experiments carried out in New Jersey and New Hampshire,[8] a calibrated tone producing apparatus was available whereby frequencies of musical atmospherics, as observed by ear, could be closely followed. It was found that in addition to tones, which could be considered as individual tweeks, there appeared at times a slight, almost unbroken resonance quality in the static. This resonance was

[10] A. H. Taylor and E. C. Hulburt, "Propagation of Radio Waves," *Phys. Rev.*, *27*, p. 189, February, 1926.

always quite obscure, which may account for its escaping observation in previous work. It appeared to consist of a band of frequencies, the midpoint of which could usually be determined with an accuracy of approximately ± 50 c.p.s. The resonance was usually observed during the evening and morning twilight periods when the damping of tweeks was high, and appeared to be closely connected with the tweeks themselves, although ordinarily showing a somewhat higher frequency. During the hours of total darkness the resonance was either absent or obscured by tweeks. At evening, resonance sometimes appeared at sunset or a short time before. Usually the first highly damped tweeks were observed at about the same time. In the early morning the resonance was observed sometimes several minutes after the last tweek.

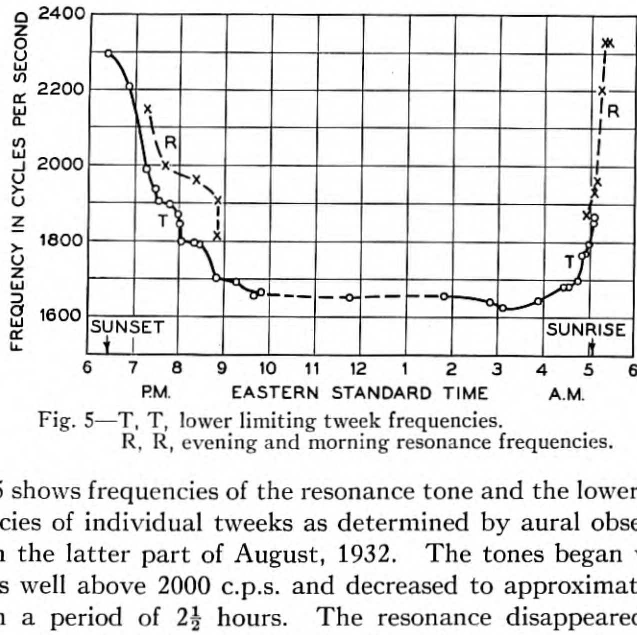


Fig. 5—T, T, lower limiting tweek frequencies.
R, R, evening and morning resonance frequencies.

Fig. 5 shows frequencies of the resonance tone and the lower limiting frequencies of individual tweeks as determined by aural observations made in the latter part of August, 1932. The tones began with frequencies well above 2000 c.p.s. and decreased to approximately 1650 c.p.s. in a period of $2\frac{1}{2}$ hours. The resonance disappeared as the tweeks approached the usual night intensity. Approximately 1/2 hour before sunrise the resonance reappeared and a rapid frequency increase began. The last definite tweek observed in the morning was still under 2000 c.p.s., although the resonance rose well above this frequency before disappearing. In approximate figures, the effective reflecting surface for audio frequencies is indicated by the data of Fig. 5 to be located at an altitude of 61 km. at sunset and to rise to 88.5 km. in a period of $2\frac{1}{2}$ hours. Half an hour before sunrise the indicated altitude is 87 km. and at 15 minutes after sunrise it has returned to 61 km.

It is possible that aural frequency observations result in erroneous determinations because of the rapid reduction in frequency which occurs during a tweek. If the damping is not excessive, the ear distinguishes the low frequencies of the tweek and thereon establishes the tonal characteristic. If the damping is great the lower frequencies may be reduced below audibility while the ear may distinguish the higher or intermediate frequencies as possessing tonal quality and thereon may base its estimation of frequency. Judging from the observations of resonance, where the sound may be almost continuous, it appears likely that these frequency determinations are of fair accuracy.

Observations have been made at various times to determine the time of appearance of the first and last tweeks of the night-time period. Fig. 6 shows the time of first tweek to be quite variable, extending from

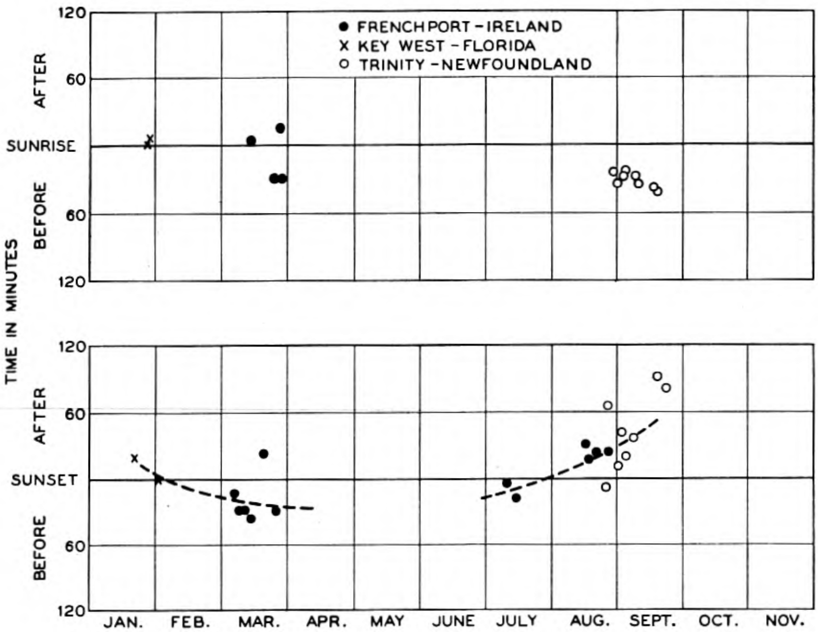


Fig. 6—Observations of first and last tweeks of night-time periods.

approximately 1/2 hour before sunset to $1\frac{1}{2}$ hours after. The time of the last tweek varies from 40 minutes before sunrise to a few minutes after sunrise. The points obtained in Florida differ somewhat from those obtained in Newfoundland and Ireland, possibly because of the difference of latitude. Since the Florida observation point lies approximately 24° south of the latter locations, it follows that here the interval between the time of incidence of the sun's rays at the position assumed for the damping region and actual sunrise is somewhat less than at the northern observation points. However, a seasonal effect may be responsible as is indicated by the dotted curve in Fig. 6.

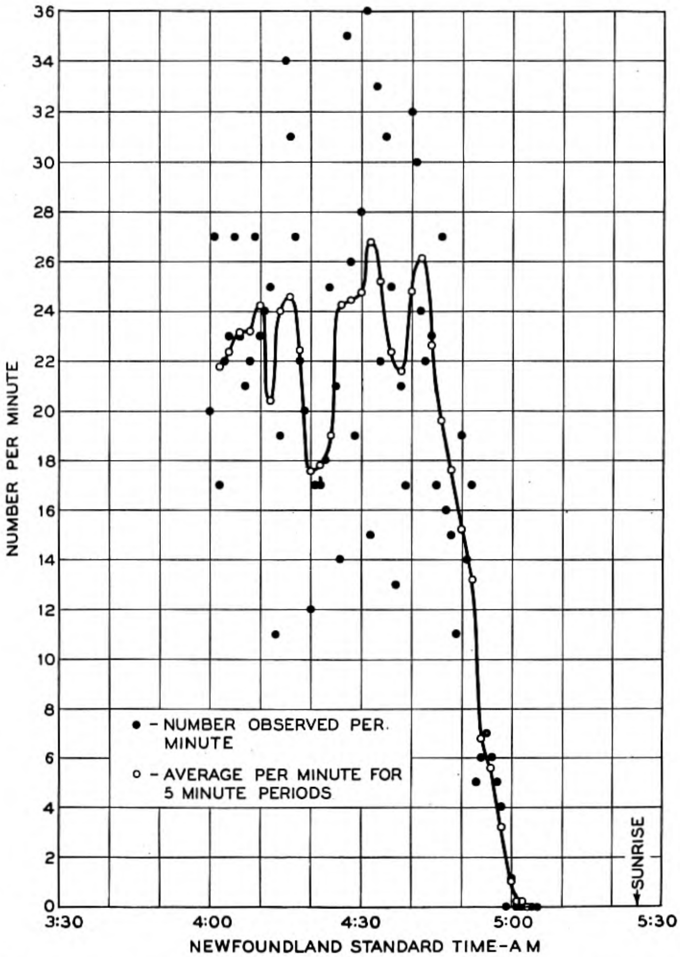


Fig. 7—Rate of occurrence of tweeks. Data taken during a period of high intensity.

There is a distinct seasonal variation in tweek numbers, the rate being consistently high during the summer and low during the winter and early spring—following approximately the variations in non-musical atmospherics. At times in the summer, tweeks have been observed to occur at rates exceeding 50 per minute while during the winter as few as one or two in five-minute periods is not unusual. A night completely free from tweeks has not been observed at any of our experimental locations. Fig. 7 shows results of a summer tweek count when the rate was high. This curve illustrates well the rapid variations which may occur during the morning twilight period.

Swish and Related Musical Atmospherics

Swishes observed in Newfoundland have been described as, "Musical sounds, such as made by thin whips when lashed through the air." [1] They are ordinarily distinctly musical in character, the frequency varying sometimes downward and at other times upward. At times upward and downward progressions are observed simultaneously. During the Newfoundland observations, the frequencies lay usually between 700 and 2000 c.p.s., but the individual tones in most cases did not exceed an octave in variation. The duration of these earlier observed swishes varied from approximately 1/4 second to more than a second. In Ireland swishes of the same nature were observed, but a more usual type was longer and much clearer in tone. These swishes were audible from 1/2 second to possibly 4 seconds and covered a frequency range from well below 800 to above 4000 c.p.s. To the ear the frequency appeared to progress steadily with perhaps a slight lingering near the termination of the descending variety.

While in the earlier Newfoundland observations the swish usually appeared to be accompanied by a rushing sound, later work disclosed many nearly clear whistling tones which may be identified as the "long whistlers" reported by other observers. These sometimes swept upward or downward through the entire voice range and at other times varied only through the range between approximately 3000 and 4000 c.p.s. On a few occasions the whistles have been observed to hesitate and warble slightly before disappearing. Series of swishes have been observed following each other with almost perfectly regular spacing of a few seconds, the train persisting on occasion for as long as a few minutes. Some of these trains have successively increased in intensity, terminating abruptly while other trains have reduced gradually until submerged in the usual static. In addition to the distinctly musical tones, swishes have been heard in which the rushing or hissing sound is prominent while the tone may be nearly or entirely absent. Our observations have shown these often to appear during periods when the whistling tones are frequent, to correspond approximately to the length of the whistles and at times to appear in regularly spaced trains.

Many observations have indicated a relation between swishes and the quasi-musical sound in the range between 500 and 1500 c.p.s., which in an earlier paper has been called "intermediate frequency noise." [1] Frequently this noise is first observed as a subdued jumble of hollow rustling or murmuring sounds. It often increases regularly in intensity for some time, after which faint swishes may begin to appear in the same frequency range. The swishes may increase in intensity and length, eventually submerging the murmuring sound. Occasionally the murmuring has continued for a short time after the swishes have reduced in amplitude or have disappeared. As a general rule the murmuring is not audibly prominent although it seems to be rather continuous in character. As a result it may considerably increase the atmospheric intensity in the intermediate voice range.

On a few occasions musical high frequencies similar in general character to the murmuring have been observed. This sound appears as a continual chirping or jingling in the vicinity of 3200 c.p.s. The amplitude is usually low and the duration short. Like the murmuring sound, it appears to accompany periods during which swishes are present, and probably is composed of large numbers of short, overlapping, high-frequency swishes.

These types of atmospherics appear to have no connection with the time of day, or with local weather conditions and there is no indication of any correlation with the time of year. During some periods they have been observed frequently during days and nights for possibly 48 hours or longer. They have been found at times to persist steadily through the early morning, bridging the transition period when the more common forms of atmospherics rapidly change character. At times several weeks of daily observation have passed with practically no appearance of swishes or related sounds.

During periods of prominent swishes the variation of intensity is usually gradual with maxima and minima spaced at irregular intervals of possibly a few minutes. At maxima, the swish may approximate the intensity of the usual audio night-time atmospherics. The intensity which swishes may attain is evidenced by their occasional observation without use of amplifying apparatus. A twelve-mile telegraph line free from power interference has been found a satisfactory antenna, and with a telephone receiver between the line and earth, swishes of remarkable clearness have been observed. Tweeks have been heard with the same equipment.

In the short time during which the sound recording apparatus was available in Ireland, swishes were very infrequent with the exception of one day when all swishes were of the descending frequency type.

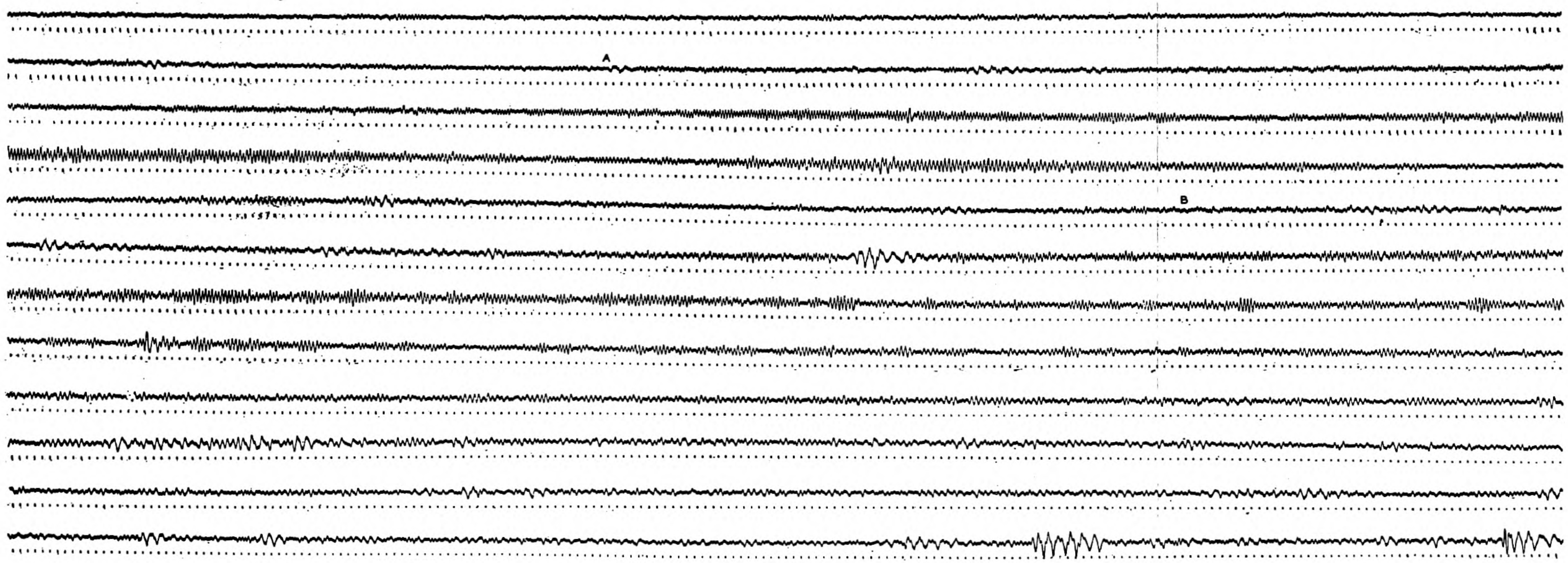


Fig. 8—Oscillogram of overlapping pair of swishes. A and B denote visible beginnings of the respective wave trains. Timing impulse frequency, 1000 c.p.s.

These swishes were unusual in that they appeared in overlapping pairs. Three minutes of record was obtained containing seven swish pairs. A representative oscillogram, shown in Fig. 8, is a record of 2.4 seconds,

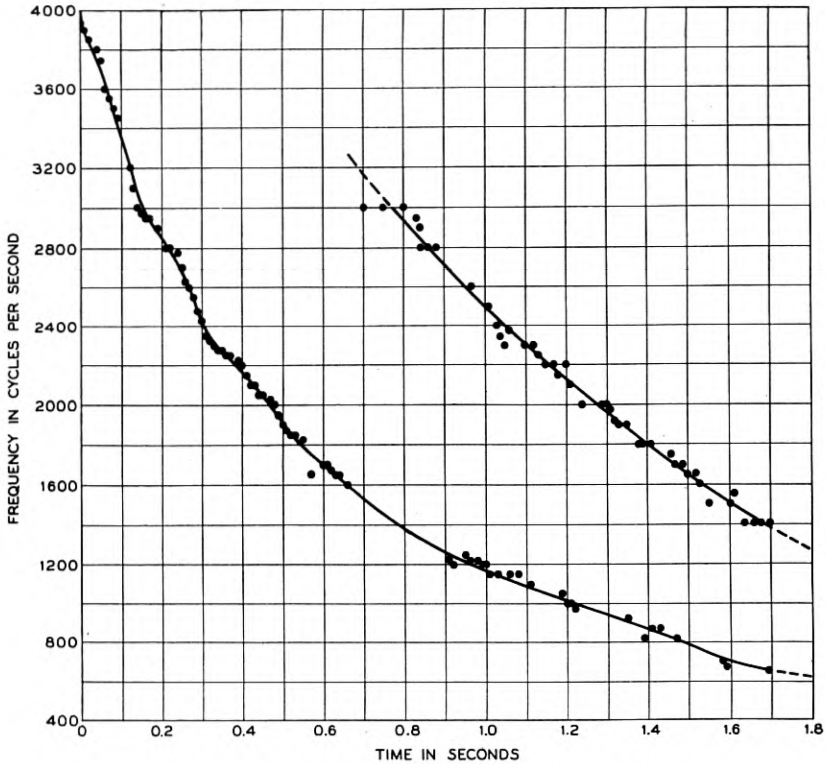


Fig. 9—Frequency curve of the swish pair shown in Fig. 8.

containing all that could be identified as a swish pair. The points "*A*" and "*B*" denote the visible starts of the first and second swishes respectively. Filters used during the recording of this oscillogram account for the absence of frequencies above 3000 c.p.s. and below 600 c.p.s. The frequency variation of this swish pair with time is shown in the curve of Fig. 9.

Eckersley [11] has reported observations of descending whistling tones following static crashes after a quiet period of a few seconds. During the New Hampshire observations this phenomenon was observed frequently. The swishes were observed to follow certain distinctive static crashes. This type of disturbance consisted of low and inter-

[11] T. L. Eckersley, "Radio Echoes and Magnetic Storms," *Nature, 122,* p. 768, November, 1928.

mediate impulses, persisting for a fraction of a second, accompanied by an unusually intense frying sound, indicating a predominance of high frequencies. At no time did this type of disturbance appear to possess marked tonal quality. Each impulse was followed by a quiet period after which a swish occurred. During several periods when the static was sufficiently intermittent, the interval between the beginning of the static impulse and the beginning of the swish was timed. Approximately 70 observations were made, the shortest period recorded being 1.2 seconds and the longest, accurately determined, 3.0 seconds. Many ranged between 2.5 and 2.8 seconds. No consistent progression of the length of this swish lag was observed although at certain times a predominance of either long or short periods existed. Later work indicated the long and short periods to be about equally divided between night and day.

During one night of the New Hampshire work an auroral arc appeared extending from northwest to northeast. Near the northwest end of the arc frequent flashes occurred, but these were too obscure for any details to be made out. A similar but much weaker flashing was observed to the southwest. At times the flashes appeared to extend along the horizon from northwest to southwest. By visual observation while listening to the atmospherics, it was found that nearly every flash coincided with a static crash possessing the prominent frying sound. These crashes were in most cases followed by swishes, usually of the descending variety, although occasionally a short ascending whistle occurred simultaneously with the start of the descending swish.

According to information supplied by the United States Weather Bureau, no lightning storms occurring during this period lay in the direction where flashes were observed to be concentrated and no storms were reported as near as 100 miles to our observation point. The Weather Bureau supplies the information that, under favorable reflecting conditions, lightning flashes might be seen 40 miles, but could not be seen 100 miles. It therefore appears reasonable to suppose that the flashes observed were of auroral origin. A report supplied by the United States Magnetic Observatory at Tucson, Arizona shows a magnetic storm beginning August 27. Through the following days the disturbance gradually reduced, reaching a low level on September 1. Our observations show the swish intensity to be high from the evening of August 30, when observations began, to September 1. Through September 1 and up to the termination of the test on the morning of September 2, the swish intensity appeared to be reducing although occasional high intensity periods occurred. These and earlier data of like nature obtained by us and others indicate a correlation between swish and magnetic disturbances. The accepted connection between auroral and magnetic field variations might justify a supposition that auroræ and whistling tones may be directly related as indicated by the New Hampshire observations. An assumption that the tones originate at the altitudes usually occupied by auroral displays might lead to an explanation of the apparent absence of marked diurnal variations in the swish tones. The observed correlation between certain atmospheric crashes and the subsequent swishes appears to indicate either dependence of the latter on the former or origin of the two from a common source of energy. The first assumption points to multiple reflection or dispersion phenomena which produce either ascending or descending tones. The time lag between the static impulse and the following swish would indicate either a low velocity or the traversing of a great distance. In either case, low attenuation is indicated by the long duration of some tones. It appears possible that the two radiations may result from sequential events occurring in the upper atmosphere by means of which non-musical as well as musical atmospherics are produced. Assuming an emission of energy which persists more or less steadily over a period comparable with the duration of a swish, it is possible to account for the approximately uniform amplitude of a swish without the necessity of assuming a very low damping.

It is suggested that swishes may be related to the occasionally observed phenomenon of swinging and flashing auroral beams. In this case it appears necessary to consider a cyclic process in the behavior of the aurora which would account for the time lag between the radiation of an initial static disturbance and the following varying tone. The varying tones might be produced by energy radiated from swinging beams resonating within the space separating beams or in the space between a beam and a stationary reflecting layer.

It might be possible for standing waves to occur within a beam, variations in the length or other constants of the path producing the varying tones.

A correlation between swish and auroral phenomena is indicated in statements by witnesses of auroral displays. Professor Chapman [12] reviews the testimony of many observers who have witnessed auroral displays at extremely low altitudes. Some attest to having stood within the glow and to having heard, directly from the atmosphere, disturbances accompanying the visible phenomena. Some of their sound descriptions follow:

[12] Prof. S. Chapman, "Audibility and Lowermost Altitude of Aurora," *Nature, 127*, p. 341, March 7, 1931.

"Quite audible swishing, crackling, rushing sounds"
"A crackling so fine it resembled a hiss"
"Similar to escaping steam, or air escaping from a tire"
"Much like the swinging of an air hose with escaping air"
"The noise of swishing similar to a lash of a whip being drawn through the air"
"Likened to a flock of birds flying close to one's head"

Some of these phrases coincide with those used by us in describing swishes. Certainly the correlation of sound descriptions is remarkable.

Dr. J. Leon Williams,[13] an observer of auroræ, comments on the sounds thus: ". . . On several occasions I have heard the swishing sound. The sound accompanies only a certain type of auroral display. I have never heard this sound except when those tall, waving columns, with tops reaching nearly to the zenith were moving across the sky. . . . When these tall sweeping columns die down the sound, according to my experience, disappears."

Consideration has been given to the likelihood of swishes or other appreciable audio frequency disturbances being produced by meteors. Lindemann and Dobson [14] estimate the energy liberation of an average meteor to exceed 3 kilowatts during the glowing period, and Skellet [15] states that a meteor may throw out an ionized trail extending laterally to a distance of a few kilometers. It has appeared advisable to search for magnetic disturbances which might show tonal qualities by resonance between the meteoric trail and some established reflecting surface. During two nights atmospherics were received with an audio-frequency amplifier and a loop antenna, located at a point in New Jersey. Observation of twenty-nine meteors, including six which could be classified as quite bright, disclosed no correlation with the sounds of audio-frequency atmospherics.

Some Theories of Musical Atmospherics

In a paper entitled "Whistling Tones from the Earth" Barkhausen [16] describes observations made during the World War on an atmospheric, which appears to have been the same as the descending swish heard by us.

He states, "During the war amplifiers were used extensively on both sides of the front in order to listen in on enemy communications. . . . At certain times a very remarkable whistling note is heard in

[13] "The Sound of the Aurora," *Literary Digest*, *112*, p. 28, February 20, 1932.
[14] Prof. F. A. Lindemann and G. M. B. Dobson, "Theory of Meteors," *Proc. Roy. Soc. Lond.*, *102*, p. 411, 1923.
[15] Skellet, "Effect of Meteors," *Phys. Rev.*, *37*, p. 1668, 1931.
[16] H. Barkhausen, loc. cit.

the telephone. So far as it can be expressed in letters the tone sounded about like *pēou*.[17] From the physical viewpoint, it was an oscillation of approximately constant amplitude, but of very rapidly changing frequency . . . beginning with the highest audible tones, passing through the entire scale and becoming inaudible with the lowest tones. . . . The entire process lasted almost a full second."

Barkhausen presents two possible explanations for these sounds. The first assumes the presence of a reflecting layer in the upper atmosphere. An electromagnetic impulse originating at the earth's surface arrives at a distant receiver first over the direct path and then from reflections in the order 1, 2, 3, to n. Such a series of reflections would result in a wave train of rapidly diminishing frequency becoming asymptotic to a value dependent upon the height of the reflector.

The second of Barkhausen's theories depends upon ionic refraction in the Heaviside layer, resulting in the breaking up of an impulse into its component frequencies and a delay in the transmission of the lower frequencies with respect to the higher. It gives a rate of frequency progression which varies with distance and with the refractive index of the medium.

Eckersley[18] in a paper on "Musical Atmospheric Disturbances" discusses apparently the same type of atmospherics. As an experimental background he notes frequent observations of audio-frequency disturbances received over large radio antennas. He states: "These (tones) have a very peculiar character: the pitch of the note invariably starts above audibility, often with a click, and then rapidly decreases, finally ending up with a low note of more or less constant frequency which may be of the order of 300 to 1000 a second.

"The duration . . . varies very considerably; at times it may be a very small fraction of a second, and at others it may be even 1/5 of a second." He observes that they are infrequent in morning, increasing throughout the day and reaching a maximum during the night. He develops a theory based on ionic refraction to account for these disturbances.

It appears that in these latter observations both swishes and tweeks were heard, but were not recognized as distinct phenomena. Such an error might be attributed to the irregularities of response which are common in the ordinary telephone receiver.

Barkhausen's first theory fails to explain swishes because of their upward as well as downward progression, long duration and frequency range. The theory, as previously pointed out, is adaptable to the

[17] *Pēou* slowly pronounced in a whisper excellently portrays a descending swish accompanied by the rushing sound.

[18] T. L. Eckersley, *Phil. Mag.*, *49*, p. 1250, 1925.

explanation of tweeks. It does not appear probable that either Barkhausen's or Eckersley's refraction theory properly explains the tweek because of its lower limiting frequency of approximately 1600 c.p.s. It seems more than mere coincidence that this frequency is in the range that the multiple reflection theory predicts. Any theory adequately explaining the swishes or long whistlers should account not only for long duration and apparently constant amplitude but for upward as well as downward progression and freedom from diurnal changes in tonal qualities.

Acknowledgments

The authors wish to acknowledge their indebtedness to Mr. A. M. Curtis and Dr. W. S. Gorton for valuable advice, and to Messrs. J. F. Wentz, A. B. Newell and E. W. Waters who through long hours have worked patiently with us in procuring the data upon which this article is based.

Certain Factors Limiting the Volume Efficiency of Repeatered Telephone Circuits

By LEONARD GLADSTONE ABRAHAM

Vacuum tube amplifiers are now regularly built into long distance telephone circuits where required to maintain their volume efficiency. Consequently, the overall volume efficiency of these circuits no longer depends to any important extent on the loss per unit length of the line wires. Instead, the efficiency is controlled by certain factors which, before amplifiers were introduced, had negligible effect. Among these factors are echo, singing or "near singing," and crosstalk. The stability of the lines and amplifiers also becomes very important.

This paper sets forth the methods now in use in the Bell System for computing the highest volume efficiencies at which telephone circuits may be worked without causing echo, singing or crosstalk effects to become too serious. The matter of making proper allowance for the normal variability of the circuits is also included. Specific references are made to various sources of published data which permit the methods to be applied to obtain practical working figures for cable circuits. The fundamentals, however, are also applicable to open-wire circuits.

THE excellence of transmission over a toll telephone circuit is determined by its overall volume efficiency (including the effect of variations from time to time), by distortion of the waves, by various delay effects and by the masking effect of noise. The term "net loss" [1] is commonly used to more specifically designate the overall volume efficiency as limited by the factors which will be discussed herein. It is equal to the total loss introduced by the toll lines and all associated apparatus minus the total gain introduced by all of the amplifiers. In the United States the net loss is usually given for the single frequency of 1,000 cycles and is expressed in decibels.

To avoid producing an undue amount of echo, singing (or near singing), or crosstalk in repeatered circuits, the net loss must be kept above certain minimum figures. The net loss which safely meets requirements for echo, singing and crosstalk after making due allowance for transmission variations in the circuit is called the "minimum working net loss." This paper discusses the methods used in the Bell System for predetermining the minimum working net losses of telephone circuits, particularly those in cable, for which references to published data are made which will enable telephone transmission engineers to readily carry out the required computations.

A telephone circuit may be used for terminal business only (i.e.,

[1] The net loss of a circuit is the insertion loss of the circuit between 600-ohm impedances.

only for calls between the two cities at which it terminates) or for through business (i.e., the circuit may be connected at one or both ends to circuits to other cities). Evidently in the case of circuits used for this second purpose consideration must be given to various combinations of circuits which may be connected together, as dealt with in the paper entitled "General Switching Plan for Telephone Toll Service" by H. S. Osborne (*B. S. T. J.*, Vol. IX, p. 429, July, 1930). Also, the working out of such a plan involves various compromises. While in working out a general transmission plan, consideration must be given to the fact that a given through circuit sometimes appears in one connection and sometimes in another, there is little difference between the computation of the minimum working net loss of a single link connection and the computation for some particular assumed combination of through circuits into a multi-link connection. The discussion which follows is written as if applying particularly to terminal circuits. However, the reader may take the methods as practically applying to a long built-up connection.

The method of determining the echo limitation is to determine the minimum echo net loss [2] and then to add an allowance for variations to determine the minimum working echo net loss. In the case of singing and crosstalk, however, the minimum working net losses are determined directly, allowance for variations being made, respectively, in the singing margin required under average conditions and in the average amount of crosstalk considered allowable. After the minimum working echo, singing and crosstalk net losses have been computed separately, the largest one of the three values is taken as the minimum working net loss of the circuit.

The echo, singing and crosstalk limitations and the normal variations are considered in detail in what follows:

Echoes

In the telephone art, the term "echo" [3] is applied to more or less faithful repetition of the conversation to which the talker or listener is a party, which reaches him through some path other than the sidetone path or the main channel of communication. If the delay of the echo is sufficient, a distinct repetition of the sound is heard which produces a sensation similar to the one usually associated with the word echo in common parlance. If the delay is very small the echo tends to merge with the sidetone or direct transmission.

[2] The minimum echo (singing, crosstalk) net loss is the smallest net loss at which a circuit, free of variations, is satisfactory with respect to echoes (singing, crosstalk).

[3] See "Telephone Transmission Over Long Cable Circuits," by A. B. Clark. (*Jour. A. I. E. E.*, January, 1923, and *Bell Sys. Tech. Jour.*, January, 1923.)

Talker echo is echo heard by the talker due to his own speech and listener echo is echo heard by the listener due to the far-end subscriber's speech. The principal effect of talker echo is to annoy and disturb the talking subscriber and perhaps to delay the conversation, but it is possible to continue talking, if necessary, despite this echo. Listener echo on the other hand may actually reduce the intelligibility but, in this case, also, the annoyance may be a considerable factor. However, the listener echo is usually less objectionable than talker echo (in circuits designed in accordance with the Bell System practice) and the following discussion will be limited to talker echo.

For a given circuit net loss and terminal return loss,[4] the absolute volume of talker echo varies with the talker volume. When there is a very long delay in a circuit, the talker echo comes back effectively separated from the outgoing speech and is objectionable if the volume of the echo is too large as compared to the circuit noise and room noise (and to some extent, perhaps, the volume of speech from the far end of the circuit). For shorter delays, the sidetone speech currents in the subscriber's set mask the echo so that it is less objectionable and the amount of masking increases as the delay decreases. In any case, the talker echo is objectionable when its volume (determined by the speech volume and the loss in the echo path) becomes too great compared to the combined masking effect of the total noise and the sidetone volume, with due regard for the fact that the sidetone currents precede the echoes.

Circuits Without Echo Suppressors

Inasmuch as the degradation of a circuit by echoes is subjective, the limitations which they place on circuit design must ultimately rest on experiments with talkers. The curve marked "No Echo Suppressor" [5] on Fig. 1 [6] shows an experimental curve of the smallest permissible net loss in an echo path for satisfactory talker echo conditions. This was obtained with typical sidetone subsets on short loops, and with typical noise conditions. It is used to find the mini-

[4] The return loss expressed in decibels between any two impedances Z_1 and Z_2 is $20 \log_{10} \left| \frac{Z_1 + Z_2}{Z_1 - Z_2} \right|$. The return loss of a repeater section or circuit, etc., is assumed to mean the return loss between that repeater section or circuit, etc., and the network circuit normally used to balance it. The terminal return loss is the return loss of the terminal switching trunk, loop and subset.

[5] The other curves on Fig. 1 were obtained at a different time and under slightly different noise, etc., conditions from those under which the upper curve was obtained.

[6] The exact effect of an echo of very short delay is not known. Such an echo will tend to increase the sidetone and thus mask any echoes of longer delay which may be present. However, in order to obtain a continuous computation method and because very short echoes are not very important in computing minimum net losses, the curves on Fig. 1 are drawn down to zero as shown. This matter and other matters in connection with echoes are being investigated further.

mum echo net loss of a four-wire cable circuit as follows: Assume a trial net loss and compute the loss in the echo path by adding the loss from the toll switchboard to the point where the echo is reflected

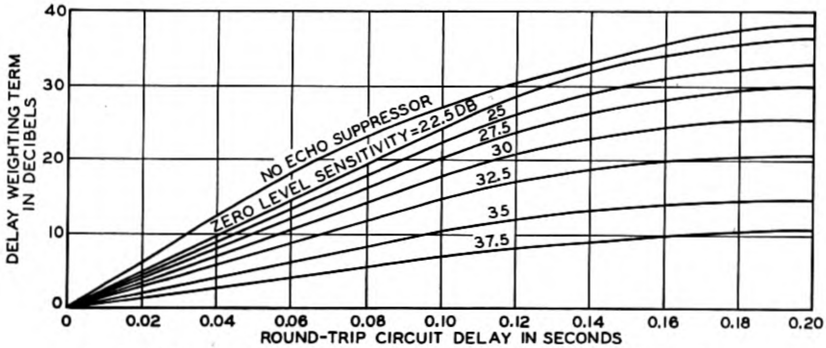


Fig. 1—Talker echo delay terms for 4-wire circuits—sidetone subsets.

back, the terminal return loss (assumed 6 db for echo computations in the Bell System) and the loss from that point back to the toll switchboard. From this total, subtract the "delay weighting term" from Fig. 1 for the corresponding round trip delay. If the resulting weighted echo path loss is greater than or equal to zero, the circuit will be satisfactory from an echo standpoint at this net loss without variations.

In the case of two-wire circuits, the echo limitations are similar to those for four-wire circuits except that echoes are also returned from intermediate points in the circuit through the return paths at the repeater hybrid coils.

The general method of determining whether circuits are satisfactory from an echo standpoint has been discussed in the paper entitled "Telephone Transmission Over Long Cable Circuits" by A. B. Clark [3] and later in a paper entitled "Echo Suppressors for Long Telephone Circuits," by A. B. Clark and R. C. Mathes (*A. I. E. E. Jour.*, June, 1925). It may be outlined briefly as follows: Determine the weighted loss of each echo path by determining the actual loss from and to the toll switchboard at the talker end (including the return loss at the point where the echo is reflected back) and then subtract the "delay weighting term" corresponding to the delay of each path as obtained from the upper curve on Fig. 1. If any one of these weighted echo path losses is reduced below zero db, the echo conditions will be unsatisfactory without regard to the effect of the other paths, as outlined above. However, if all these losses are positive, it is considered that the net effect of all of the paths may be determined by adding the power ratios (less than unity for a loss greater than zero) of the individual weighted losses together and finding the equivalent weighted echo path corresponding to this sum. When this equivalent path becomes zero db (a power ratio of 1.0), the circuit (without variations) is considered to be just satisfactory from an echo standpoint.

The distribution of gains between the different repeaters in a two-wire circuit usually has an appreciable effect upon the minimum net loss which may be obtained for a given circuit. If the gain in each direction of transmission of each repeater is equal to the loss of the preceding repeater section (or is less than it by a fixed amount called the taper), it may be shown that the echo limitations computed as above are completely determined by the delays involved, the taper, the terminal return loss and by the differences between the return loss, S, and attenuation loss, L, of the repeater sections, i.e., the values of S-L.[7] The minimum echo net loss of any given two-wire circuit (for given terminal conditions), therefore, is determined by the delays, S-L, the taper and the number of repeater sections. The value of S-L which is of the greatest importance is usually that in the important echo range, i.e., about 500 to 1,500 cycles.

While the terminal return loss is taken as a fixed value (6 db) in these computations, the return loss at intermediate repeater points varies according to the structure of the line. The statistical distribution of the return losses of loaded cable circuits may be computed as outlined by Crisson.[8] It is customary to compute the return loss, S_L, at 1,000 cycles, using the distribution function $S_F = 0$ in Crisson's formulas. To determine the echo limitations, the value $S_M = S_L - 4$ is used, principally to take into account the fact that the computed values of S_L are at a single frequency.

In addition to the return loss of the bare cable facilities, the return loss of the repeating coils and other office equipment and the effect of the termination at the far end of the repeater section must be considered. These components are:

$$S_1 = S_M + 2C,$$
$$S_2 = S_C,$$
$$S_3 = S_T + 2L + 2C,$$

where S_1, S_2 and S_3 are the return losses (attenuated to the repeater),

[7] In the following, this is assumed the same for each repeater section. It may be seen that the use of S-L instead of S and L separately effectively removes one variable from computations.

[8] "Irregularities in Loaded Telephone Circuits," by George Crisson, *B. S. T. J.*, Vol. IV, and *Elec. Comm.*, Vol. 4, October, 1925. Specific values of the deviations from which S_H may be computed are given in a paper entitled "Long Distance Telephone Circuits in Cable," by A. B. Clark and H. S. Osborne.

respectively, of the bare cable, the near-end apparatus and the terminating effect at the far end of the repeater section.

C = apparatus loss at near end.

S_C = return loss of apparatus at near end.

S_T = terminating effect of repeater and apparatus at far end of repeater section.

L = loss of the line section at 1,000 cycles.

The overall return loss of the complete repeater section, S, is assumed equal to the combination of S_1, S_2 and S_3 as the sum of the corresponding power ratios.

Circuits With Echo Suppressors

When echoes would otherwise be objectionable on a circuit, it may be equipped with an echo suppressor. On a four-wire circuit equipped with an ordinary echo suppressor, the currents which are strong enough to operate the echo suppressor have their echoes suppressed. When currents are too weak to operate the suppressor, echoes will be returned, but, of course, will be much weaker than the loudest echoes on the same circuit without an echo suppressor. The echoes on the circuit with an echo suppressor will, therefore, generally be less objectionable than those on the same circuit without an echo suppressor, since those which get back to the talker are weaker in absolute volume, while the noise and sidetone volume for a given speech volume are unchanged.

The more sensitive the echo suppressor is made, the weaker the sounds will be which will just fail to operate the suppressor. Consequently, the echoes will become less objectionable as the sensitivity is increased. However, if the sensitivity is increased too much, the suppressor may be falsely operated by noise currents, either from the circuit, from room noise at the subscriber's premises which is picked up through his transmitter, or from room noise picked up through operators' sets.

The process of determining the minimum echo net loss of a circuit equipped with an echo suppressor has the following two steps: (1) determine the zero level sensitivity [9] of the echo suppressor on the circuit which is allowable with little or no false operation from noise and (2) determine the minimum net loss from experimental curves.

[9] The zero level sensitivity is defined as the amount of loss it is necessary to insert between a 600-ohm source of one milliwatt of power and the 600-ohm input of the circuit on which an echo suppressor is located in order to cause the echo suppressor to be just operated. Unless otherwise specified, this is assumed to be at 1,000 cycles.

First, determine the maximum amount of noise (including room noise and the effect of variations in net loss) which may be expected at the echo suppressor input in an appreciable number of cases. If this noise is N db above reference noise,[10] it has been determined experimentally that the local sensitivity [11] which will cause the echo suppressor to be steadily and completely operated is about (90-N) db. Providing a reasonable margin against noise operation to allow for different kinds of noise and the like, the safe local sensitivity is about (80-N) db.

The value so determined is the maximum allowable local sensitivity. From this value, the maximum allowable zero level sensitivity is obtained by adding the gain from the circuit input to the echo suppressor input under the net loss conditions for which the local sensitivity was computed. The average allowable zero level sensitivity is less than the maximum allowable zero level sensitivity by the negative variations in net loss and echo suppressor sensitivity (the negative variations are the amount by which the average loss is decreased) which may be expected. In the Bell System, average zero level sensitivities of about 31 db on toll circuits may be considered typical.

To compute the minimum net loss on a four-wire circuit, assume a trial net loss and determine the loss in the echo path as outlined above for circuits without echo suppressors. From this loss, subtract a delay weighting term from Fig. 1 for the corresponding round trip circuit delay on the proper curve. With an echo suppressor near the center of the circuit,[12] the delay weighting term is read on the curve for the average zero level sensitivity. As before, if the resulting weighted echo path loss is greater than or equal to zero, the circuit (without variations) will be satisfactory from an echo standpoint.

In general, echo suppressors on two-wire circuits have not been found desirable in the Bell System. However, a layout of considerable interest occurs when a two-wire circuit is connected in tandem with a four-wire circuit equipped with an echo suppressor. The computation of the echo limitations is approximately as outlined above

[10] Reference noise is equal to one micro-microwatt (10^{-12} watt) at 1,000 cycles or the equivalent weighted power at other frequencies or combinations of frequencies.

[11] The local sensitivity is defined as the amount of loss it is necessary to insert between a 600-ohm source of one milliwatt of power and a 600-ohm resistance across which an echo suppressor is bridged in order to cause the echo suppressor to be just operated. Unless otherwise specified, it is assumed to be at 1,000 cycles.

[12] In the Bell System, echo suppressors are generally located near the center of the circuit. If the echo suppressor were not near the center of the circuit, due allowance for the relative variations of the zero level sensitivity and the circuit net loss should be made. For example, for an echo suppressor at the end of a circuit, the zero level sensitivity as measured from the far end would be practically a maximum when the lowest net loss was obtained.

for two-wire circuits without echo suppressors, except that the delay weighting terms for all echo paths which are acted upon by the echo suppressor are determined from the curve for the proper zero level sensitivity. The paths which are not affected by the echo suppressor are all paths which do not pass through the suppressor and any paths with enough delay beyond the suppressor so that the hangover [13] is insufficient to suppress the echo. (Echoes in this latter class are normally not obtained, since the hangover is made large enough to suppress all echoes beyond the suppressor.)

Singing and Circulating Currents

Another effect which is important principally on two-wire circuits is that of singing and circulating currents. In a two-wire circuit, if the total gain around a repeater is increased sufficiently, it will become greater than the losses across the hybrid coils and singing will occur if the phase relations are right. When this occurs, the subscriber may hear the singing tone, repeaters may be overloaded, voice-operated devices on connecting circuits may be falsely operated and other circuits in the same cable may be made noisy by cross-induction.

Even when actual singing does not occur, if the loss minus the gain around a circulating path is small, the voice currents may be considerably distorted due to the feedback currents around the repeater. If the singing margin [14] becomes small, the circulating current or "near-singing" effect is quite objectionable.

In order to provide against this possibility, it has seemed desirable in the Bell System to require a 10 db singing margin [14] around the most critical repeater in any long circuit, under average conditions of temperature, regulation, net loss, etc., and with 5 db terminal return losses. (For circuits equipped with only one or two repeaters, 8 db margin is considered sufficient.) In a similar manner to that outlined above for echoes on two-wire circuits, the quantity *S-L*, the taper, and the terminal return loss are the important things in determining this singing margin. In this case, of course, the delay does not have any large effect. The value of *S-L* which is usually of the most importance is the one at about the highest frequency efficiently transmitted, since this usually tends to be the lowest value of *S-L* within that range.

The process of computation of the singing margin around a given

[13] This is the same as the "releasing time" discussed in the paper entitled "Echo Suppressors for Long Telephone Circuits" mentioned above.

[14] The singing margin is the sum of the additional gains in the two directions which may be inserted at the most critical repeater in the circuit before singing starts, under specified conditions as to the terminations, etc.

repeater is as follows: The active return loss [15] in one direction, say east of the repeater under consideration, is first obtained (Fig. 2). The passive return loss of the adjacent repeater section toward the

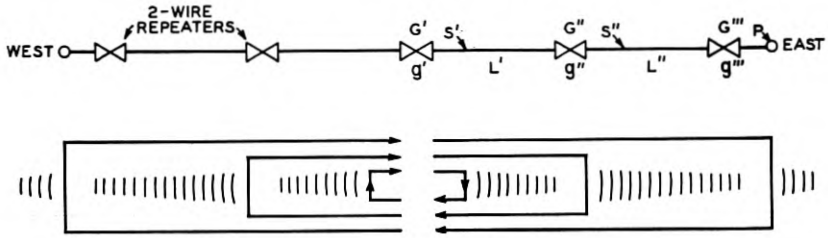


Fig. 2—Singing paths in a 2-wire circuit.

S' and S'' are passive return losses of cable sections.
P is the terminal return loss.
G', G'' and G''' are west to east gains.
g', g'' and g''' are east to west gains.

east (S') constitutes the first singing path and is determined as outlined above in considering echoes on two-wire circuits, except that the 4 db is not subtracted because singing occurs at only one frequency and because approximately the worst frequency is selected for computations. The passive return loss in the repeater section on the far side of the next repeater to the east (S'') is amplified and attenuated through the intervening gains ($G'' + g''$) and losses ($2L'$) to obtain the second component, which is $L' - G'' + S'' - g'' + L'$. Similar components are determined for all other repeater sections to the east of the one under consideration. (In the case of the circuit shown on Fig. 2, there are no more such paths.) These paths are then combined by adding the power ratios corresponding to these paths. The loss of the resultant singing path is the active return loss from the repeater under consideration with no currents returned from beyond the terminal repeater (or from the circuit terminal if there is not a terminal repeater). This active return loss is then combined with the path including the terminal repeater, viz., ($L' - G'' + L'' - G''' + P - g''' + L'' - g'' + L'$), according to the sum of their current ratios to obtain the active return loss (toward the east from the repeater in question) of the circuit in normal operating condition. (The use of current ratios rather than power ratios in this case is indicated by theoretical considerations and confirmed by experimental data.)

The active return loss toward the west from the repeater in question

[15] An active return loss is a return loss with gain inserted in the paths of one or more of the returned currents. The passive return loss is the return loss without any currents returned from beyond the adjacent repeater (or other termination if there is not a repeater there).

is then determined in a similar manner. The sum of these two active return losses minus the two-way gain of the repeater in question is approximately the singing margin around that repeater.

Whatever singing margin is obtained under average conditions, there will be certain factors tending to reduce this margin while the circuit is in normal operation. These factors include net loss variations, transmission-frequency characteristic deviations, removal of one of the normal terminations for short intervals, gain lumping due to pilot wire regulation, and slight troubles which have not yet been corrected. Because of those factors, and because of the disadvantages of near singing, 10 db singing margin under average conditions (8 db for short circuits) is believed desirable in the Bell System.

Crosstalk

Net losses may also be limited by the danger of excessive crosstalk. Far-end crosstalk from circuit 1 to circuit 2, each extending from *A* to *B*, is crosstalk which manifests itself at the *B* end of circuit 2 from the speech of the subscriber at *A* on circuit 1. Near-end crosstalk from the same talker may manifest itself at *A* on circuit 2.

From a general standpoint, the crosstalk volume should be so low that no subscriber can understand what any other subscriber says on another circuit. This is desirable from the standpoint of preserving secrecy and also from the standpoint of the annoyance which may be caused by unwanted speech currents.

The assumed limitation on circuits from a crosstalk volume standpoint is that a subscriber shall have only a very small chance of hearing understandable crosstalk. This chance is determined by the distributions of the crosstalk couplings, the room noise and circuit noise, the terminal losses, the talker volumes on other circuits, and the natures of the talkers and listeners. Present data indicate that the chance of a subscriber hearing understandable crosstalk is very small in the case of two-wire cable circuits if the crosstalk conditions are such that there is not more than about one chance in 100 that any one or more of the couplings between the disturbed circuit and the various disturbing circuits shall exceed 1,000 crosstalk units (60 db loss). Further investigations of this matter and other questions in connection with crosstalk are being made.

Crosstalk in cable circuits may be either within-quad or between-quad crosstalk. Crosstalk within the quad may be phantom-to-side, side-to-phantom or side-to-side and may be divided into office crosstalk and cable crosstalk.[16] The office crosstalk is due to capacitance

[16] Specific values of the various sources of crosstalk are given in a paper entitled "Long Distance Telephone Circuits in Cable," by A. B. Clark and H. S. Osborne, *B. S. T. J.*, Vol. XI, Oct., 1932.

unbalance in the office wiring and to repeating coils, repeaters, and other office apparatus.

The crosstalk in the cable outside the office is due to loading coil unbalance, series resistance unbalance, and capacitance unbalance. Crosstalk between different quads is normally due almost entirely to capacitance unbalance.

When the complete repeater sections have been installed, cross-connection of the circuits at certain repeater points is generally used to reduce the overall crosstalk between circuits. In the case of two-wire circuits, this cross-connection consists of breaking up all phantom-to-side and side-to-side combinations in a given quad at each repeater station, and the system is designed to make it improbable that any two of these circuits will ever be in the same quad again. In the case of four-wire circuits, this cross-connection is resorted to only at the ends of each pilot wire regulator section.

The method of computing the crosstalk limitations of a given cable circuit is as follows: Determine the r.m.s. (root mean square) within-quad crosstalk coupling per loading section by adding together the r.m.s. crosstalk coupling due to capacitance unbalance, resistance unbalance and loading coil unbalance as the r.s.s. (root sum square) of the parts expressed in crosstalk units. From this, get the r.m.s. unamplified crosstalk coupling per repeater section by properly attenuating the crosstalk coupling from each loading section. The attenuation in each case equals the loss from the output of the repeater transmitting into the disturbing circuit (in that repeater section) to the point of crosstalk coupling plus the loss from this point to the input of the repeater receiving from the disturbed circuit. The total r.m.s. within-quad crosstalk coupling per repeater section is the r.s.s. of the crosstalk coupling from each of the loading sections and from the office. The between-quad crosstalk coupling per repeater section is obtained in a similar manner.

In the case of near-end crosstalk on two-wire circuits, the unamplified crosstalk coupling so determined is then amplified or attenuated by the gains or losses from the transmitting terminal of the disturbing circuit to the repeater section in question and then to the receiving terminal of the disturbed circuit. Next, the r.s.s. of this crosstalk coupling and the between-quad crosstalk coupling from the same disturbing circuit in other repeater sections is obtained. The probability of this total crosstalk coupling exceeding 1,000 units is then determined, making due allowance for the variations of net loss. For near-end crosstalk, in a circuit without variations, the probability that 1,000 units of crosstalk will be exceeded when the total r.m.s. crosstalk coupling [17] is "x" crosstalk units is approximately

$$P_n = e^{-k^2} \text{ where } k = \frac{1{,}000}{x}.$$

An approximate method of allowing for circuit variations is to consider a circuit with variations equivalent to a circuit without variations with a net loss smaller than the average net loss of the former circuit by one-quarter of the variations; i.e., if the variations are $\pm\ V$ db, the value of k to be used in the above formula is

$$k = \frac{1{,}000}{x} 10^{-(V/80)}.$$

Fig. 3 shows the value of P_n plotted against k.

When these probabilities have been determined for all circuits having a similar within-quad exposure to the circuit under consideration, the total probability of the crosstalk coupling exceeding 1,000 units from any circuit may be determined and is approximately the sum of the probabilities of excessive crosstalk coupling from each of the disturbing circuits. (The probability of excessive crosstalk from circuits not having within-quad exposures is considered negligible.) When this probability is .01, the circuit is considered just satisfactory from a crosstalk standpoint.

Far-end crosstalk coupling is computed in a similar manner, using the probability relations applying to far-end crosstalk and four-wire circuits, which are somewhat different from those applying to near-end crosstalk and two-wire circuits. In this case, the probability of exceeding 1,000 units of crosstalk when the r.m.s. total crosstalk is "x" units is approximately

$$P_f = 1 - \frac{2}{\sqrt{\pi}} \int_0^{k/\sqrt{2}} e^{-t^2} dt, \text{ where } k = \frac{1{,}000}{x},$$

or with variations of $\pm\ V$ db,

$$k = \frac{1{,}000}{x} 10^{-(V/80)}.$$

Fig. 3 shows P_f plotted against k.

VARIATIONS

When the minimum net loss at which a circuit will be satisfactory has been determined, or when the minimum working net loss is com-

[17] The ratio of the average near-end crosstalk to the r.m.s. near-end crosstalk is about $\sqrt{\pi}/2$. The similar ratio for far-end crosstalk is $\sqrt{2}/\sqrt{\pi}$.

puted directly, it is necessary to make an allowance for the fact that the circuit will vary from time to time. The principal variations in an unregulated cable circuit are caused by temperature changes. In

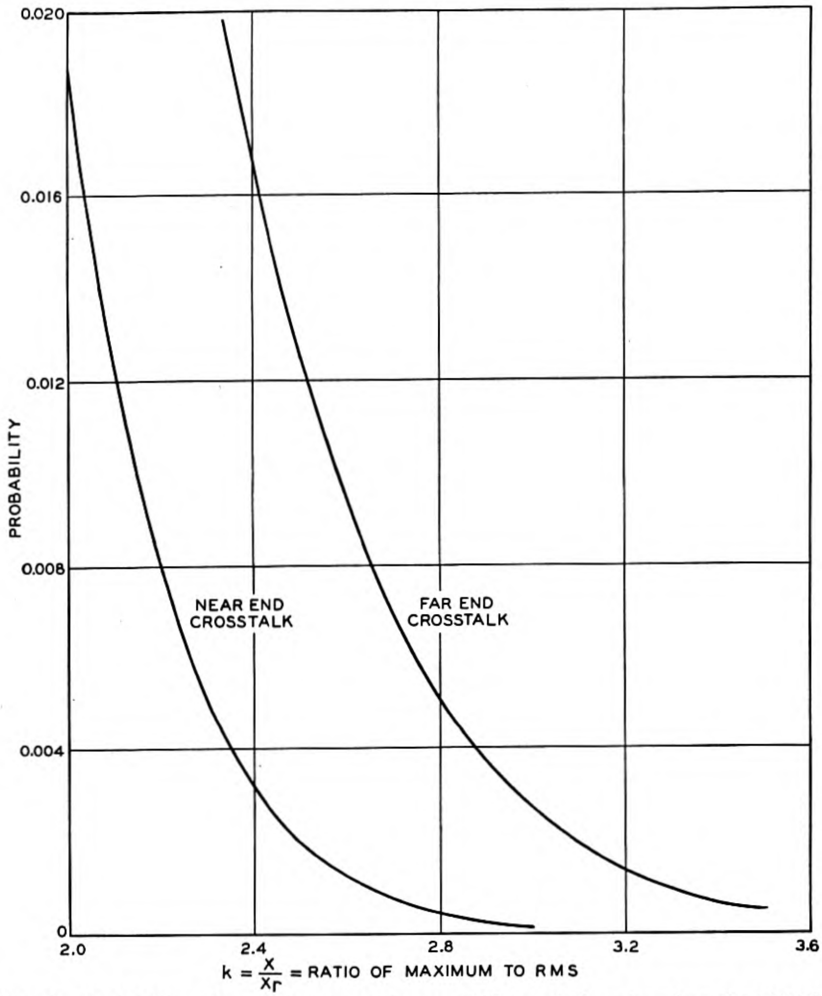


Fig. 3—Probability of exceeding a maximum of X crosstalk units when the R.M.S. is X_r crosstalk units.

a 1,000-mile circuit of 19-gauge H-44-25 four-wire facilities in aerial cable in the northeastern part of the United States, for example, a variation at 1,000 cycles of about $\pm$ 55 db from the average would be normally expected due to temperature variations throughout a year. About 35 to 45 per cent of this would occasionally occur in one day while in shorter circuits somewhat higher percentages might be encountered. In underground cable about one third as much would be encountered in a year, but very little would normally occur in one day since the rate of change is small.

In order to take care of these large variations a system of pilot wire regulation is used. The following discusses this system in some detail in order to show what the residual variations are. This system consists of a pilot wire extending through the cable whose circuits are to be regulated, each pilot wire being perhaps 100 to 150 miles in length. An automatic mechanism measures the d-c. resistance of this pilot wire frequently, and makes occasional adjustments of the gain of the regulating repeaters. In the case of the four-wire facilities referred to above, these adjustments are made in approximately .5 db steps at 1,000 cycles, and other suitable adjustments are made at other frequencies.

This pilot wire is placed in the four-wire part of the cable (it is usually obtained by compositing a four-wire circuit) and therefore has very closely the same temperature variation as the four-wire pairs. The position in the cable of the two directions of transmission of four-wire circuits is reversed [18] at the center of each repeater section, so it is possible to regulate both directions of transmission from a pilot wire in either group without serious error. Since the two-wire circuits are comparatively short, have generally smaller variations in decibels per mile than four-wire circuits, and usually have an average position in the cable, there is no serious error in regulating these from the same pilot wire.

Due to the finite steps in which these regulators operate, there is a residual variation which is approximately $\pm .25$ db per regulating repeater. In addition, there may be a certain amount of lag in the operation of these regulators, because of the fact that it is desirable to prevent excessively frequent operation of these devices, and perhaps partly because of mechanical backlash. To prevent hunting it is necessary to make the adjustment in the pilot wire regulator somewhat smaller than the adjustment which would be necessary to make all the variation due to this cause a random matter. In other words, when the temperature is changing in a given direction in many repeater sections, for example early in the morning, the adjustment at each of the pilot wire regulators is slightly behind what it theoretically should be for the pilot wire resistance obtaining at that time. This results in a directly additive effect in all regulating repeaters in a given circuit during certain times of day. By careful design and routine

[18] This assumes concentric segregation which is generally used.

maintenance, it is possible to reduce this effect to about $\pm .03$ db per regulating repeater. Other regulation inaccuracies, including imperfections in the design and manufacture of regulating networks and departure of individual pairs from average characteristics, may introduce an additional error of about $\pm .1$ db per regulator, this effect being more or less random, however.

In addition to the residual effects of temperature changes there are variations in the net losses of the circuits due to repeater battery changes and humidity changes. The repeater batteries are usually held to fairly narrow limits and vacuum tubes are tested regularly for emission. The expected change in repeater gain due to an "A" battery change of $\pm .5$ volt is about $\pm .2$ db and for a "B" battery change of $\pm$ 5.0 volts is about $\pm .25$ db.

In office cabling and in the switchboard multiple at the terminals of the circuit there may be a considerable amount of variation due to changes in the humidity. This has been largely taken care of by improvements in the type of cable used (cellulose acetate) and by keeping the lengths of office cable as short as possible. However, a residual variation of about $\pm .5$ db may be expected, a considerable part of which is due to switchboard multiple.

If the number of repeaters in a circuit is "n" and the number of regulators is "r," the total variations are considered to be about

$$V_1 = \pm \sqrt{(.5 + .03r)^2 + (.25)^2 r + (.1)^2 r + (.2)^2 n + (.25)^2 n}.$$

These items are allowances respectively for humidity variations, regulator lag, finite regulator steps, other regulator errors, "A" battery changes and "B" battery changes. Rearranging the equation,

$$V_1 = \pm \sqrt{.25 + .1025r + .0009r^2 + .1025n}.$$

In addition to this variation, the probability that the average net loss of a given circuit is not exactly as specified must be considered, so the variation from the specified value is considered to be about $\sqrt{2}\,V_1$ or

$$V_2 = \pm \sqrt{.5 + .205r + .0018r^2 + .205n}.$$

Assuming that each of the individual variations from the various sources has an equal probability throughout its range, the probability that the overall variation V_2 will be exceeded is about .085, and the probability that the average variation in the two directions of transmission (which is of considerable interest in singing or echo computations) will exceed this is still smaller.

Example

As an example of the general procedure in specifying satisfactory net losses for terminal circuits, consider a 500-mile 19-gauge H-44-25 four-wire cable circuit not equipped with echo suppressors. From the information in the paper by A. B. Clark and H. S. Osborne referred to above:

1. The minimum echo net loss is about 4.5 db.
2. The transmission variations are about ± 2 db.
3. Therefore, the minimum working echo net loss is about 6.5 db.
4. The minimum working crosstalk net loss is about 6.6 db.
5. The minimum working singing net loss is about 0 db.
6. Therefore, the minimum working net loss of the circuit is about 6.6 db.

It will, therefore, be satisfactory to specify 6.6 db with normal variations of ± 2.0 db for the circuit in question.

Abstracts of Technical Articles from Bell System Sources.

The Effect of Temperature on the Emission of Electron Field Currents from Tungsten and Molybdenum.[1] A. J. AHEARN. Electron field currents from the central portion of long molybdenum and tungsten filaments about 2.7×10^{-3} cm. in diameter have been studied. The field currents were first made stable to about 5 per cent by long-continued conditioning treatments of temperature and high voltage under high vacuum conditions. Thermionic emission measurements gave the values 4.32 and 4.58 volts for the work function of the molybdenum and tungsten, respectively, in good agreement with the accepted values for the clean metals. Emission measurements were then made at fields varying from about 5×10^5 volts/cm. to about 1×10^6 volts/cm. and at temperatures varying from 300° K. to about 2000° K. Down to about 1600° K. the thermionic currents completely masked the field currents. Thermionic emission values below 1600° K. were obtained by extrapolation. Thus the field currents at the lower temperatures were separated from the thermionic currents. Where necessary, corrections were made for the decrease in the voltage gradient accompanying the thermal expansion of the filament. The field currents were found to be independent of temperature to within 5 per cent from 300° K. to 1400° K. At temperatures higher than 1400° K. the data are consistent with the assumption that the current consists of a thermionic current plus a current which is independent of temperature. However, because of the exponential change of thermionic current with temperature a small effect of temperature on the field current could not be distinguished at temperatures higher than 1400° K. From the theory of Fowler and Nordheim, β, a factor introduced by surface irregularities, is found to be 120 for the tungsten cathode and 47 for the molybdenum one. Thus for tungsten, Houston's theory of the temperature effect is in approximate agreement with the negative results of these experiments.

Measurement of Transmission Loss Through Partition Walls.[2] E. H. BEDELL and K. D. SWARTZEL, JR. This paper reviews the theory and describes the method used at Bell Telephone Laboratories of measuring the transmission loss through partition walls. The partition to be

[1] *Phys. Rev.*, August 15, 1933.
[2] *Jour. Acous. Soc. Amer.*, July, 1933.

tested is built into an opening between two adjacent but structurally isolated rooms. A loud speaker acts as a source of sound in one room and a portion of the sound energy is transmitted into the second room through the test partition. The transmission loss is taken as

$$TL = L_1 - L_2 - 0 \log_{10} (\alpha_2/A),$$

where L_1 and L_2 are the intensity levels in the source and test room respectively, expressed in db, α_2 is the absorption in the test room and A is the area of the partition. The levels L_1 and L_2 are measured and plotted with a moving coil microphone and an automatic level recorder, and a beat frequency oscillator is used as a source of tone so that the frequency may be varied continuously. Measurements with a continuous variation in frequency enable resonances in the partition to be much more easily and quickly detected than is possible when measurements are made at discrete frequency intervals. Both pure and frequency modulated tones have been used for the measurements. Results of measurements on a few partitions are given.

The Optical Behavior of the Ground for Short Radio Waves.[3] C. B. Feldman. The rôle of the ground in radio transmission is first considered generally. In short-wave propagation taking place via the Kennelly-Heaviside layer only the ground in the vicinity of the antennas is involved, and its effect may be included in antenna directivity. The utility of so ascribing the ground effect exclusively to the terminals of a radio circuit rests on the applicability of simple wave reflection theory in which the distance between the terminals does not appear. For this purpose reflection equations, similar to Fresnel's equations for a nonconducting dielectric, are employed with a complex index of refraction.

The paper describes experiments undertaken to determine the limits of applicability of these optical reflection equations and discusses the results. Particular emphasis is placed on the identification of direct and reflected waves. The existence of a surface wave, foreign to simple reflection theory, is recognized with vertical antennas, when the incident wave is not sufficiently plane. At angles of incidence between grazing and the pseudo-Brewster value the requirements of planeness are severe. The relation of optics to Sommerfeld's theory is discussed. The experiments include tests made with the aid of an airplane.

For short-wave communication via the Kennelly-Heaviside layer, use of the modified Fresnel equations is shown to be justified. These

[3] *Proc. I.R.E.*, June, 1933.

equations fail only at substantially grazing incidence and then merge into the Sommerfeld ground wave solution. The ground effect is always to discriminate against radiation or reception at very low angles.

Two methods of determining the electrical constants of the ground are described. One comprises measurements of the elliptical polarization of the ground wave, and is based on Sommerfeld's propagation theory. The other is a method of measuring, at radio frequencies, the conductivity and dielectric constant of samples of ground removed from the natural state. Suitable agreement between the two methods is found if the nonuniformity and stratification of natural ground is considered. The sample method is also used to determine the conductivity of ocean water.

On Minimum Audible Sound Fields.[4] L. J. SIVIAN and S. D. WHITE. The minimum audible field (M.A.F.) has been determined from data taken on 14 ears over the frequency range from 100 to 15,000 c.p.s. The observer is placed in a sound field which is substantially that of a plane progressive wave, facing the source and listening monaurally. The M.A.F. is expressed as the intensity of the free field, measured prior to the insertion of the observer. Similar data are presented for binaural hearing, over the range from 60 to 15,000 c.p.s., obtained with 13 observers. At 1000 c.p.s. the average M.A.F. observed is 1.9×10^{-16} watts per cm.^2, corresponding to a pressure 71 db below 1 bar. Included are data showing how the M.A.F. varies with the observer's azimuth relative to the wave front. Another type of threshold data refers to minimum audible pressures (M.A.P.) as measured at the observer's ear drum. The differences obviously to be expected between M.A.F. and M.A.P. values are due to wave motion in the ear canal and to diffraction caused by the head. The M.A.F. data are discussed in relation to the M.A.P. determinations from several sources. Some possible causes of difference between the two, which are due to experimental procedure and may add to the causes already mentioned, are pointed out.

Naturally-Occurring Ash Constituents of Cotton.[5] A. C. WALKER and M. H. QUELL. Precise information on the inorganic ash constituents which are deposited in cotton fibres during growth, and on the changes which occur in these constituents when cotton is washed with distilled water or aqueous solutions, is desirable as an aid in understanding many of the properties of this important industrial fibre. In a

[4] *Jour. Acous. Soc. Amer.*, April, 1933.
[5] *Journal of the Textile Institute*, March, 1933.

previous paper reference was made to laboratory experiments in which raw (untreated) cotton was washed with distilled water and various aqueous solutions, and sufficient analytical data were given to show the effects of changes in the ash constituents upon the electrical properties of the cotton.

It is the purpose of this paper to present a discussion of the analytical data obtained in these experiments, together with a possible distribution of the ash constituents as salts occurring in the raw cotton. This distribution is based upon a somewhat unusual consideration of the analytical data. It will be shown that ionic interchange occurs when cotton is washed in aqueous salt solutions, the principal effect being the replacement of Mg^{++} in the cotton by Ca^{++} from $CaSO_4$ solutions used in washing, or the reverse if the solutions are $MgSO_4$. Although these analytical data were secured in an investigation of the electrical properties of cotton, they are the subject of a more general discussion in this paper, since it is possible that they may be of service in the study of other properties of cotton or other forms of cellulose.

Influence of Ash Constituents on the Electrical Conduction of Cotton.[6] A. C. Walker and M. H. Quell. It has been shown that the electrical properties of textiles, such as cotton, silk, wool, and cellulose acetate silk, depend to a remarkable extent upon their moisture contents and chemical compositions. In addition, these properties have been considered to depend upon water-soluble, electrolytic impurities present in the fibres, since the insulation resistance of untreated cotton has been improved very greatly by water washing.

Evidence will be presented in this paper to show that the improvement in d-c. insulation resistance of cotton, secured by washing, is accompanied by a reduction in the inorganic ash content from about 1 per cent of the dry cotton weight to a value generally less than 0.3 per cent. Data will be given to show that the water-soluble salts present in raw cotton, which constitute about 70 per cent of the ash weight, are principally potassium and sodium salts, and their removal by washing is accompanied by an improvement of between 50 and 100 fold in the insulation resistance. Since these salts are largely inorganic electrolytes, this improvement in resistance is termed *electrolytic*. A *total* improvement of between 150 and 200 fold can be secured if the washed cotton is dried under certain conditions. The difference between *electrolytic* and *total* improvement is due to changes in the moisture-adsorbing properties of the textile resulting from the manner of drying, and this difference, largely reversible by subsequent ex-

[6] *Journal of the Textile Institute*, March, 1933.

posure of the cotton to high atmospheric humidities, is termed *transient* improvement.

The effects of ash constituents, other than Na and K, on the insulating properties of cotton are small, and these effects are difficult to evaluate, since they are masked by the effect of atmospheric humidity.

In this investigation, primary consideration has been given to cotton since it is the most economical material available for use in telephone apparatus insulation, and the improvements in electrical properties secured by water-washing have led to its substitution for silk to a large extent in the telephone industry.

Contributors to this Issue

LEONARD GLADSTONE ABRAHAM, B.S., 1922, M.S., 1923, University of Illinois. American Telephone and Telegraph Company, Department of Development and Research, 1923–. Mr. Abraham has been engaged in transmission development work on toll telephone systems.

E. M. BOARDMAN. Attended Parsons College and Iowa University, 1923–26; Physics Department, Yale University, 1927–28. Bell Telephone Laboratories, 1929–. Mr. Boardman has been engaged in studies of submarine cable interference.

E. T. BURTON, A.B., 1920, M.A., 1924, Indiana University. Lieutenant, U. S. Engineers' Corps, 1917–18. Research Department, Western Electric Company, and Bell Telephone Laboratories, 1920–. Mr. Burton has been engaged in research connected with low frequency and carrier frequency signaling systems.

R. F. DAVIS, B.E.E., Cornell University, 1921. American Telephone and Telegraph Company, Department of Operation and Engineering, 1921–. Mr. Davis' work has been largely concerned with the electrical protection of communications circuits and with the electrical coordination of such circuits with power transmission and distribution circuits.

JOHN G. FERGUSON, B.Sc., University of California, 1915; M.Sc., 1916; Research Assistant in Physics, 1915–16. Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Mr. Ferguson's work has been in connection with the development of methods of electrical measurement.

HARVEY FLETCHER, B.Sc., Brigham Young University, 1907; Ph.D., University of Chicago, 1911; Instructor of Physics, Brigham Young University, 1907–08, and University of Chicago, 1909–10; Professor, Brigham Young University, 1911–16. Engineering Department, Western Electric Company, 1916–25; Bell Telephone Laboratories, 1925–. As Acoustical Research Director, Dr. Fletcher is in charge of investigations in the fields of speech and audition.

H. R. Huntley, B.S., University of Wisconsin, 1921. Engineering Department, Wisconsin Telephone Company, 1917–30; Department of Operation and Engineering, American Telephone and Telegraph Company, 1930–. Mr. Huntley's work has been principally concerned with transmission and inductive coordination matters.

W. A. Munson, A.B., University of California, Southern Branch, 1927. Bell Telephone Laboratories, 1927–. Mr. Munson has been working on speech and hearing problems.

Albert C. Walker, B.S., Massachusetts Institute of Technology, 1918; Ph.D., Yale University, 1923. Bell Telephone Laboratories, 1923–. Dr. Walker has been engaged in developing and applying methods of improving the electrical properties of textile insulation and methods for the inspection control of commercially purified textiles for telephone apparatus.