The Bell System Technical Journal

April, 1934

The Carbon Microphone: An Account of Some Researches Bearing on Its Action *

By F. S. GOUCHER

A great variety of speculations in regard to the physics of microphonic action has arisen because of the complexity of behavior when current passes through a so-called "loose contact" which forms the essential element in a carbon microphone. Technical difficulties arising from the minuteness of the contact forces and movements between contacts when in a sensitive microphonic state have retarded the establishment of a quantitative theory.

Recent studies of carbon contacts have led to a satisfactory picture of the nature of such contacts and their mode of operation when strained, both from the elastic and the electrical point of view. The surfaces of the carbon particles are microscopically rough and when two such surfaces are brought together under the action of compressional forces, both the number of hills in intimate contact and the contact area between hills vary through deformations which are primarily elastic. Changes in electrical resistance under strain are consistent with the assumption that current passes through the regions in intimate contact.

Introduction

FEW electrical devices are as widely used as the "carbon microphone" and few have given rise to as much speculation in regard to their mode of action. That the problem has proved elusive is shown by the fact that in Bell Telephone Laboratories it has been regarded as perennial. However, recent researches have thrown a considerable amount of light upon it and it therefore seems fitting to bring before you this evening a brief survey of the subject and an account of some of the latest experimental work.

The widespread use of the "carbon microphone"—it is employed almost exclusively throughout the world in commercial telephone service—is due primarily to its unique property of being its own amplifier. In converting acoustical into electrical waves, it magnifies the energy about one thousandfold. Other microphones, such as the condenser or electromagnetic type, are unable to do this and so require separate amplifiers when used in practice. For this reason, it seems unlikely that the carbon microphone will be supplanted in the near future for at least the great bulk of telephone work.

* Presented before the Franklin Institute, March 2, 1933. Published in the *Journal of the Franklin Institute*, April, 1934.

The essential element of this device is what has come to be called the "loose contact"—or, as its name implies, a contact between two conductive solids, metals as well as carbons, held together with small forces. The ability of "loose contacts" to transmit speech was discovered independently by Emile Berliner in this country and Professor D. E. Hughes in England. Following Hughes' discovery, Mr. Spottiswood, the president of the British Association in 1878, described it thus: "The microphone affords another instance of the unexpected value of minute variations—in this case, electric currents; and it is remarkable that the gist of the instrument seems to be in obtaining and perfecting that which electricians have hitherto most scrupulously avoided, viz., 'loose contacts.'" Hughes applied the word "microphone" to his instrument because of its remarkable "ability to magnify weak sounds." The word itself is a revival of a term first introduced by Wheatstone in 1827 for a purely acoustical device developed to amplify weak sounds. Although originally confined to the "loose contact" type of instrument, the term microphone has more recently been used—particularly in broadcast, public address, and sound picture work—for any device which converts sound into corresponding electric currents.

Evolution of the Carbon Microphone

The story of the development of the "loose contact" type of microphone is a fascinating one and, although it is beyond the scope of this

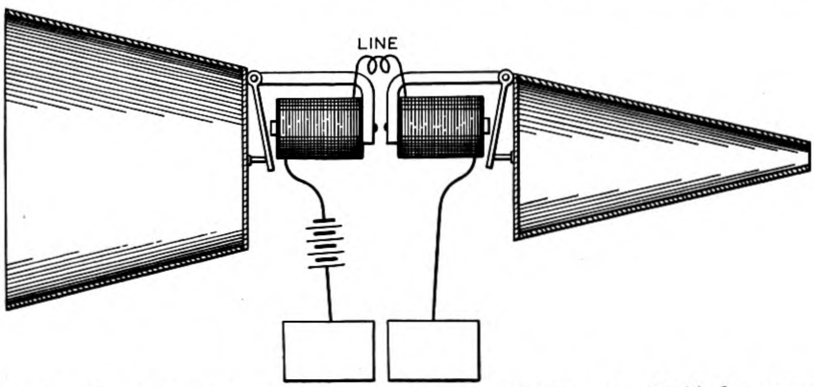

Fig. 1—Sketch, illustrating Bell's conception of the telephone, used in his first patent application of 1876.

paper,[1] I should like to refer briefly to a few of the stages in the evolution of the present day instrument. You will recall that Bell's original telephone (Fig. 1) was electromagnetic in principle and acted

[1] For a more complete account see paper by H. A. Frederick, "The Development of the Microphone," *Bell Telephone Quarterly*, July, 1931.

both as a transmitter and as a receiver. It was, however, very inefficient and Bell himself suggested that some other principle such as that of variation of electrical resistance might overcome the difficulty. He therefore devised the liquid transmitter in which a small platinum wire (Fig. 2), attached to a drumhead of gold-beaters

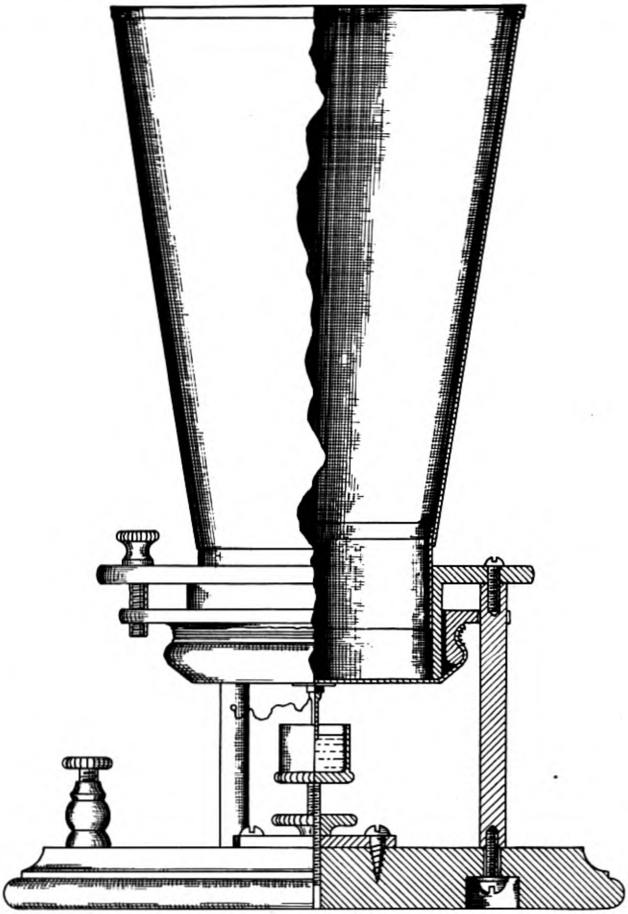

Fig. 2—Bell's liquid transmitter.

skin, is dipped into a small quantity of acidulated water in a conducting cup. The extent of the area of contact between the liquid and the wire is altered by the motion of the latter, thus altering the resistance in a continuous manner. It was with this instrument that the first complete sentence, "Mr. Watson come here—I want you," was successfully transmitted on March 10, 1876. This achievement stimulated others to work on the problem of a variable resistance element and many new devices appeared in the next few years, the most sensitive of which utilized a single loose contact, carbon in one form or another being used as the contact material.

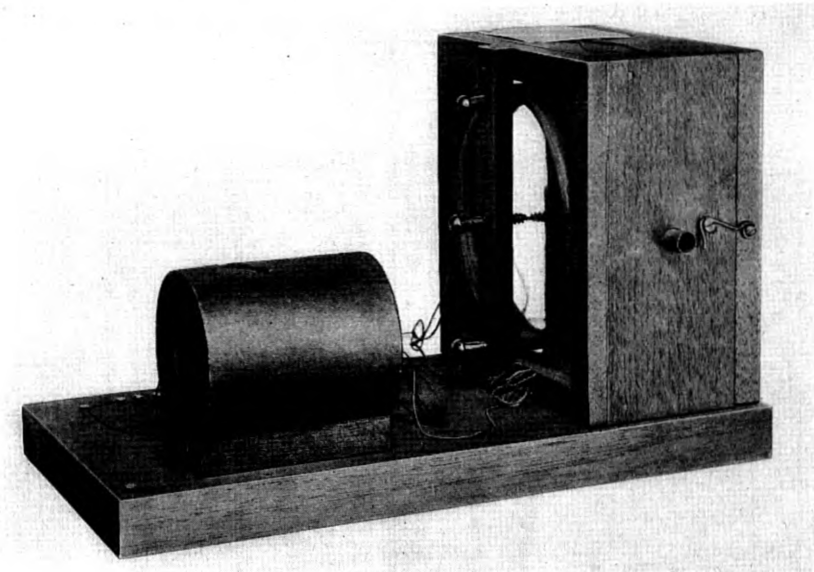

Fig. 3—Berliner's first single contact microphone, invented in 1877, employing a metal-to-metal contact.

Figure 3 shows Berliner's first successful model consisting essentially of a metal contact pressed against a metal diaphragm. This was developed later into a carbon-to-carbon contact along the same lines (Fig. 4).

Hughes, too, used metal in his first successful attempt at transmitting sounds. Only three ordinary nails were required to demonstrate the great sensitivity of loose contacts to acoustical vibrations (Fig. 5). Hughes later developed the pencil type of microphone (Fig. 6) in which carbon was used. It was the forerunner of many practical devices developed along this line.

More rugged, reliable and permanent than either of these types was the Blake transmitter shown in Fig. 7. It utilized a metal-to-carbon contact and it owed its success to the mechanical control of the contact pressure. This instrument was used for many years by the Bell System.

Then came the Hunnings or the first of the granular carbon microphones (Fig. 8), the immediate ancestor of the granular carbon type used today. Hunnings used powdered "engine coke." It carried

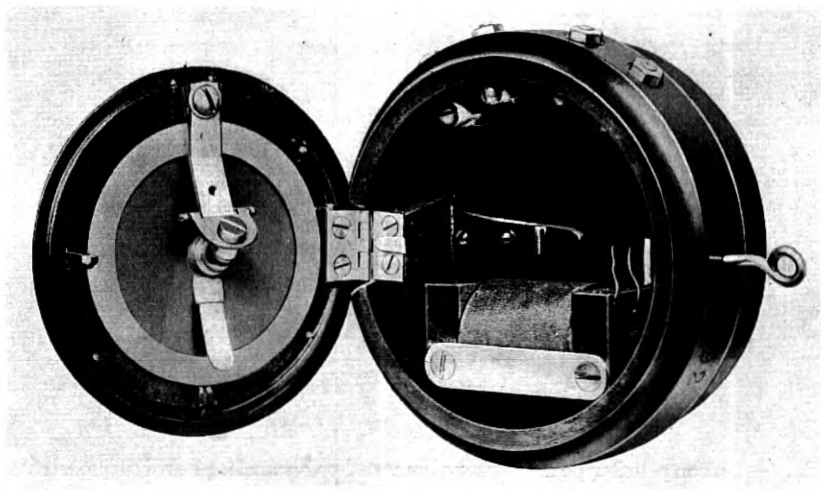

Fig. 4—Carbon-to-carbon single contact transmitter brought out in 1879 by Berliner.

more current than the Blake transmitter but it was liable to "pack" and become insensitive.

This difficulty was overcome in the design invented by White in 1890, called the solid back type (Fig. 9). Millions of these are used today in the ordinary desk-stand instrument. In this, carbon granules

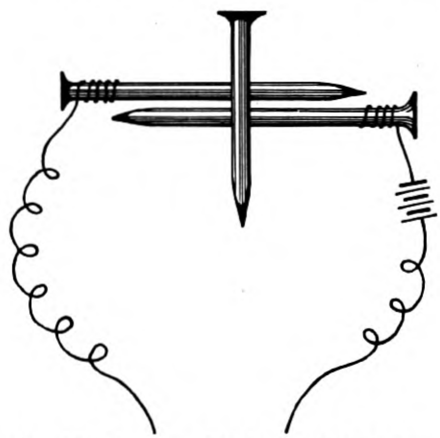

Fig. 5—Nail contacts used by Professor Hughes in 1878 to demonstrate their microphonic properties.

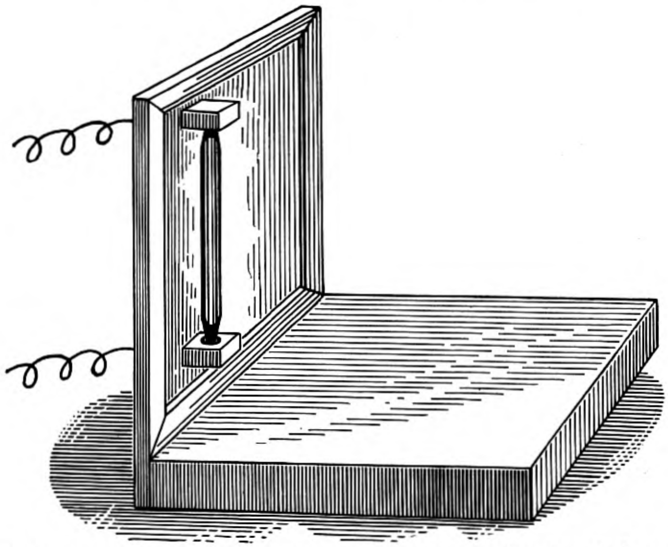

Fig. 6—Carbon pencil type microphone, mounted on a sounding board, demonstrated by Hughes in 1878.

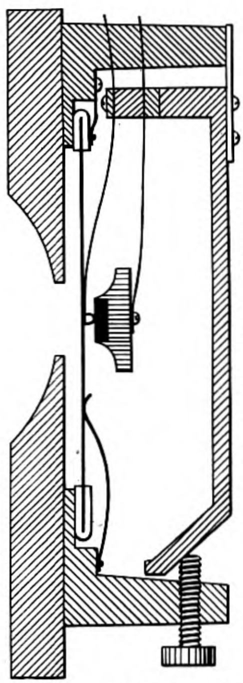

Fig. 7—The Blake transmitter using a platinum contact pressed against a carbon block.

are compressed between two polished carbon electrodes which are immersed in the granular mass in such a way that the particles have more freedom of movement than in the Hunnings instrument. This relieves excess pressure without undue packing.

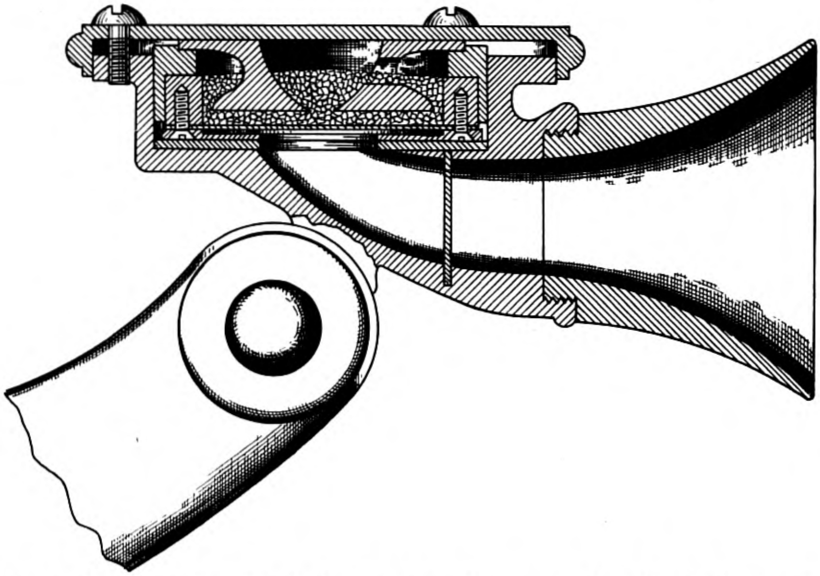

Fig. 8—Commercial model of the early Hunnings transmitter in which granular material was first used.

In Fig. 10 we have a cross-sectional view of a modern handset transmitter. This instrument, which is designed to operate in a wide variety of positions, follows the Hunnings' type in that the granular mass rests against the diaphragm but it differs from it in that the diaphragm does not act as an electrode. Both electrodes, separated by an insulating barrier, form part of the containing walls of the cell holding the carbon. This is the type which has recently been studied in detail and of which a two dimensional model is shown in Fig. 26.

The carbon used in these instruments is made by a heat treatment of anthracite coal. The particles are about 0.01 inch in size and when magnified they look just like lumps of coal taken from the domestic pile (Fig. 11).

Speculations of the Early Inventors

Part of the difficulty in elucidating the microphonic action of the "loose contact" arises because so many effects can be observed or are associated with the action that it is hard to determine which of them is essential. It is therefore not surprising that there was great diversity of opinion amongst the early inventors.

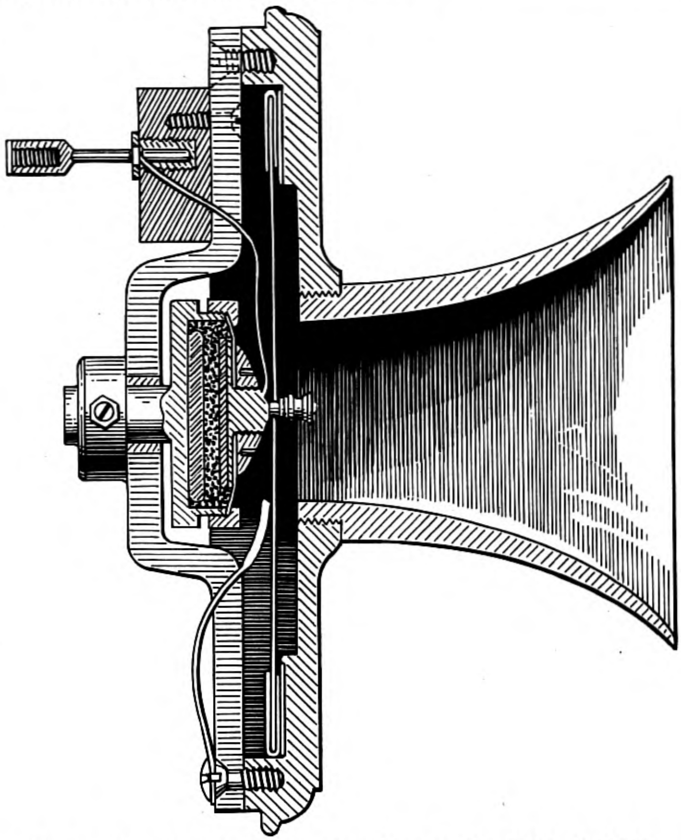

Fig. 9—The solid back transmitter invented by White in 1890.

For instance, experiment shows that contacts tend to move apart when in the act of transmitting sound. This led many, amongst them Berliner, to hold the view that an air film is necessary for microphonic action, that the current somehow passes through the film, and that the variation of the current is due to the variation of the thickness of the film. This view, however, was partly discredited by experiments showing that the moving apart was probably due to a heating of the contact through the passage of current and hence that it is not a necessary accompaniment of microphonic action.

Again, when one listens through a receiver placed in a circuit containing a "loose contact," noises are heard, especially when the voltage across the contact or microphone is large. These noises are irregular like frying or crackling. Also, if a contact be viewed under a microscope, bright spots are sometimes seen. These facts have led many to think that small arcs are always present and are responsible for microphonic action. Hughes was very much inclined to this view.

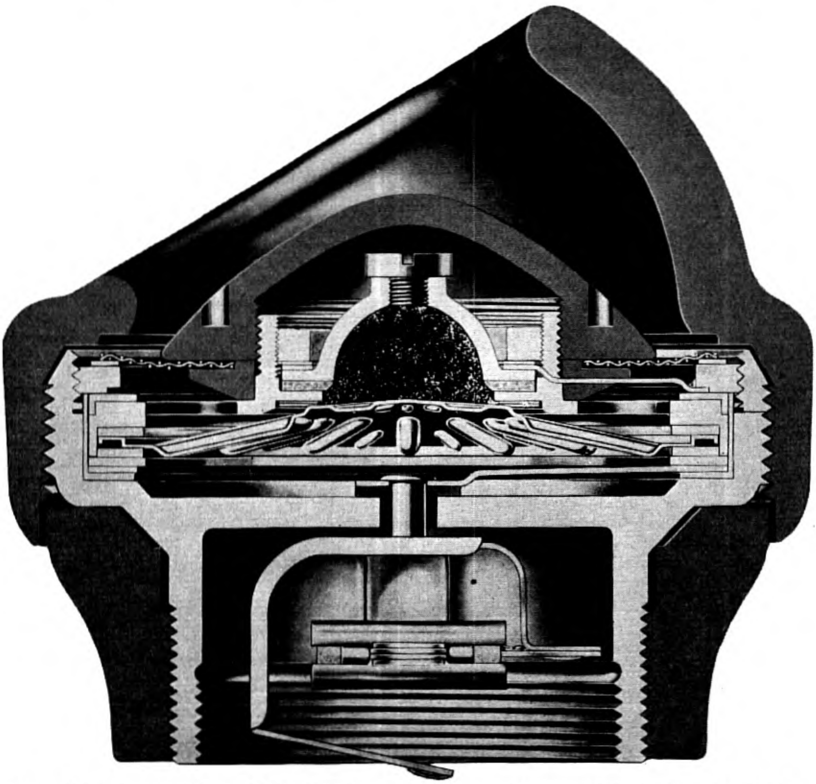

Fig. 10—Cross-section of the barrier type transmitter used in modern handset instruments.

There were reasons for supposing that the heating of the contact is a necessary factor in microphonic action. This point of view was supported by Preece, who wrote in 1893, "Indeed there are many phenomena such as hissing and humming that are clearly due to what is known as the Trevelyan effect, that is, the motion set up by expansion and contraction of bodies which are subjected to variation in temperature. This at least tends to favor the heat hypothesis as does also the fact that with continuous use some transmitters become essentially warm."

Another view was that microphonic action arises from change in resistivity of the solid carbon resulting from strain. This view was held by Edison who doubtless believed it because of the success of his microphone which was designed with the object of applying pressure variation to a solid carbon block. It failed of general

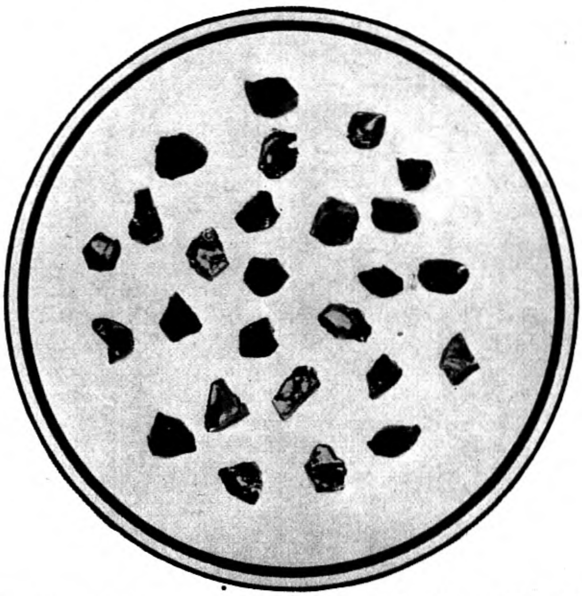

Fig. 11—Carbon granules made from anthracite coal (× 15).

acceptance because the effect of pressure on resistance, as shown by experiment, seemed definitely to be too small. It was generally considered that the Edison instrument was in fact a "loose contact" although Edison himself did not realize it.

Others of the early inventors considered the contact area to be the essential element—that is to say, the extent of surface or the number of molecules involved in intimate contact. As Professor Sylvanus Thompson expressed it in 1883, "An extremely minute motion of approach or recession may suffice to alter very greatly the number of molecules in contact. . . . Just as in a system of electric lamps in parallel arc the resistance of the system increases when the number of lamps is diminished and diminishes when the number of lamps connecting the parallel mains is increased, so it is with the molecules at the two surfaces of contact."

Recent Theories

The first attempt at a quantitative theory of microphonic action was made by Professor P. O. Pedersen in 1916.[2] He assumed that microphonic action is due to the variation of the contact area arising from the elastic deformation of the contact material by pressure. Considering the case of two elastic conducting spheres brought into contact, Pedersen assumed that the resistance is made up of two parts; viz., (1) the resistance of a conducting film having a specific resistivity differing from bulk carbon and independent of pressure, and (2) the so-called "spreading resistance" or that which is caused by the concentration of the current flow within the region of the contact area and which would exist independently of any film.

This theory results in a quantitative expression [3] for the dependence of the contact resistance on the force holding the contacts together. Pedersen tested it by experiments on carbon spheres and found reasonable agreement over a wide range of force. However a very similar expression can be obtained without postulating the existence of the high resistance film. We have merely to suppose that contact does not take place over the whole contact area owing to surface roughness (the existence of which can be observed under a microscope, especially in the case of carbon).

Dr. F. Gray of Bell Telephone Laboratories worked out an expression [4] based on this assumption which was so nearly like Pedersen's that it was difficult to discriminate between them experimentally. He assumed both that the number of microscopic hills in electrical contact increases as the contact force is increased and that the resistance per hill varies in accordance with the theory of spreading resistance as assumed by Pedersen. His equation was found to fit experimental curves remarkably well for contact forces which are relatively larger than those holding the granules together in a microphone. In the range of smaller forces, however, marked departures from theory were found, the measured value of resistance decreasing too rapidly with an increase of force. Although these departures were believed to be due at least in part to a plastic deformation of the contact material, it appeared possible that other factors come into play and may even be dominant in this region of small contact forces.

For instance, it had been demonstrated that adsorbed films of air are capable of producing a marked increase in the resistance of granular carbon contacts. This revived the air film theory as a possibility under the condition of small contact forces.

[2] *The Electrician*, Jan. 28–Feb. 4, 1916.

[3] $R = AF^{-2/3} + BF^{-1/3}$.

[4] $R = AF^{-7/9} + BF^{-1/3}$ (*Phys. Rev.*, **36**, 375, 1930).

Again there is a marked decrease in the resistance of granular carbon contacts with increase in voltage which had not been satisfactorily explained. This fact suggested amongst other possibilities that the conduction process may involve the passage of electrons across gaps of molecular dimensions in the manner of a cold point discharge. Field gradients of sufficient magnitude to extract electrons from a solid must exist in these gaps with only a fraction of a volt across the contacts. If this is the main process by which current passes between contacts, microphonic action might well be associated with a variation of the gap dimensions under strain.

Again recent work on the theoretical strength of solids had led to experimental results showing that under certain conditions solids may, without fracture, be subjected to strains greatly exceeding those heretofore obtained. This suggested the possibility that the microphonic effect of contacts might after all be associated with the straining of small junctions welded under pressure and current.

In view of the speculative nature of the situation it was clear that a new experimental attack on the problem was necessary. We have been making such an attack during the last few years and I now turn attention to some of the experimental results and the main conclusions to be drawn from them.

Recent Experimental Work

Statement of the Problem

Since the essential element in the carbon microphone is the so-called "loose contact," the first and most fundamental step toward the understanding of the physics of microphones is the solution of the problem of the "loose contact" when in its sensitive or microphonic state.

Measurements on microphones such as the handset have enabled us to specify pretty accurately the conditions under which any two granules within the structure operate when the microphone is transmitting speech or sound.

In addition to the voltage, which is limited to one volt per contact, these conditions may be stated briefly either in terms of contact forces or in terms of movements between centres of granules. When you realize how small these are—particularly the movements between centres of granules—you will, I think, not be surprised that the solution of the problem of the "loose contact" has been so long delayed.

For the condition of reasonably loud speech the diaphragm motion is about

$$1 \times 10^{-5} \text{ cm.,}$$

which is just on the limit of resolution of the highest-power microscopes. It follows from a consideration of the number of granules in series that the movement between centres of granules would not be greater than 1/10th of this, viz.,

$$1 \times 10^{-6} \text{ cm.},$$

which is in the submicroscopic range. We must, therefore, be able to control and measure movements at least as small as 10^{-7} cm.; not an easy thing to do with a "loose contact."

The contact forces are on the average somewhat less than 10 dynes when the aggregate is in the unagitated state. In the presence of acoustic waves, variable forces of several dynes are superimposed on these fixed forces. The variable forces are smaller than the fixed forces, so that the granules will on the average remain in contact throughout any reversible cycle. We have reason to believe that 10 dynes is about the maximum force which is attained at any one contact during a stress cycle. We must therefore be able to control contact forces within the range 1 to 10 dynes.

Apparatus and technique have now been developed for studying single contacts within the prescribed range of forces and displacements, and significant measurements have been made which I will now endeavor to describe to you somewhat in detail.

Single Contact Studies

Figure 12 shows the construction of one of the contact tubes used in this study.

Its essential features are shown diagrammatically in Fig. 13. The contact pieces C_1 and C_2 are fastened respectively to a movable base M and to the lower end of a helical spring made of fused quartz. The base is supported from a fixed frame by two vertical platinum wires P and two stretched springs as shown. The lower contact piece is moved by heating or cooling the platinum wires through the passage of current. In this way the contacts may be made or broken and any desired contact force applied, the measure of the force being the compression of the helical spring. The temperature of the contact is varied by surrounding the contact region with a metal cylinder S which may be heated by means of radiation from a coil of platinum wire H, the temperature within the cylinder being measured by means of a thermocouple placed near the contacts.

In practice the upper contact piece consists of a single granule fastened to the end of a platinum wire and the lower contact piece consists of a number of granules attached to a horizontal metal plate;

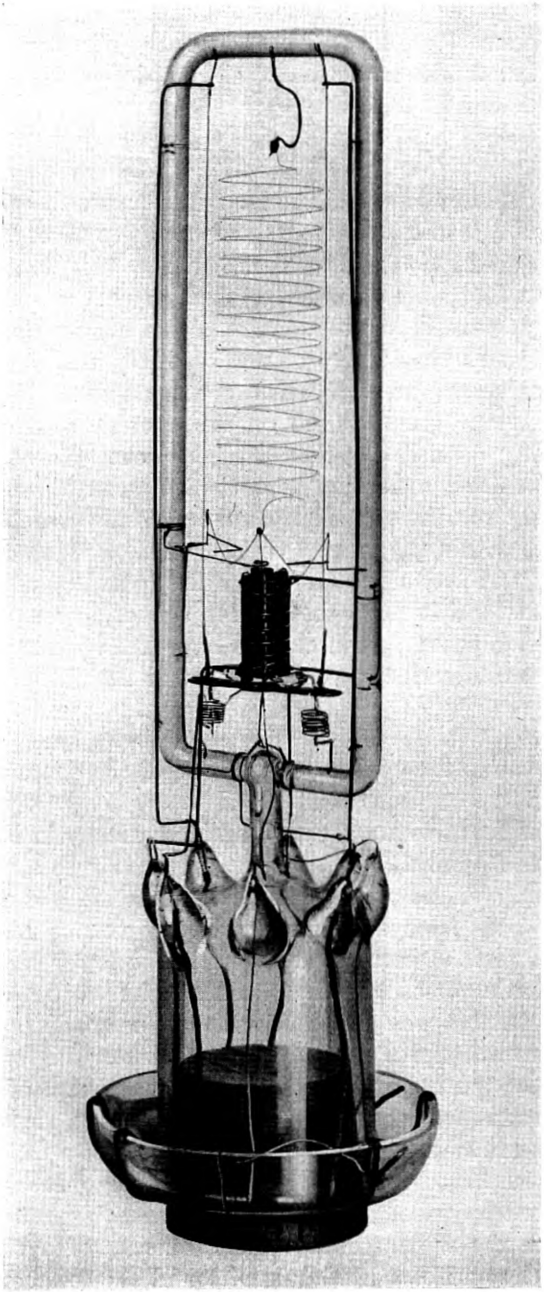

Fig. 12—Device for controlling force and temperature used in the study of single contacts.

in this way a variety of contacts can be studied with the same tube. A small hole in the metal cylinder permits of direct observation of the contacts during measurement. Figure 14 shows how the apparatus

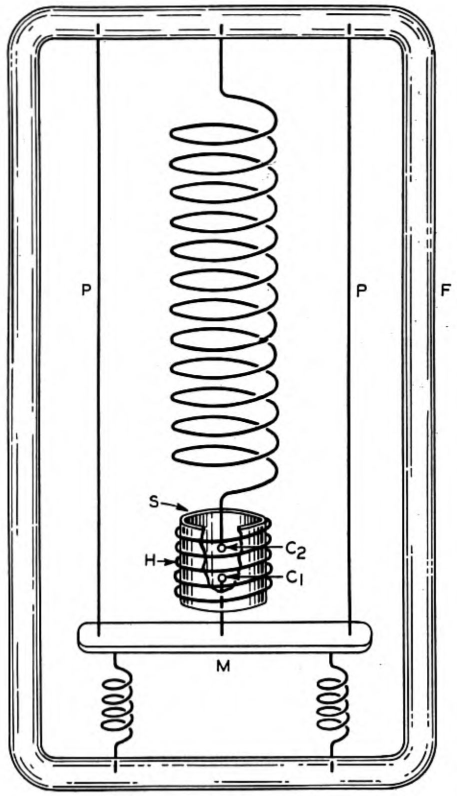

Fig. 13—Diagrammatic view of single contact device shown in Fig. 12.

was mounted in an iron cylinder on a damped suspension to protect it from acoustical and mechanical disturbance. The two microscopes were used to observe the compression of the silica spring.

We first studied the effect of voltage and temperature on contacts held together with constant forces. Reversible characteristics could in all cases be obtained for voltages up to 1 volt and for temperatures up to about 80° C.

Typical characteristics are shown on Fig. 15 in which the contact forces were of the order of 1 dyne. On the left are plotted the resistance-voltage characteristics and on the right the resistance-temperature characteristics. All of the variables are plotted for convenience on logarithmic scales.

The curves I, II and III illustrate the fact that Ohm's law is found to hold for all contacts up to about 0.1 volt and that above these values the contact resistance decreases with increase of voltage.

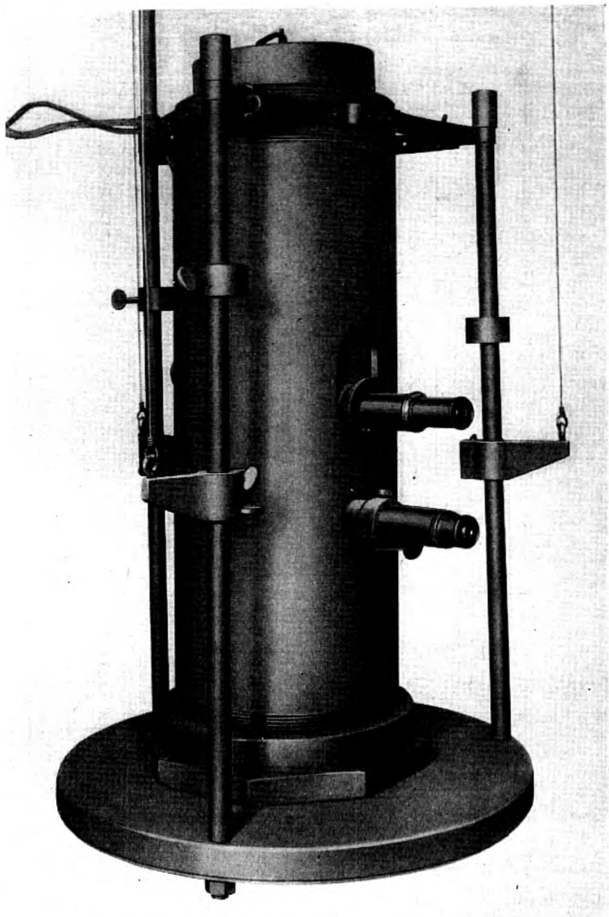

Fig. 14—The single contact device is mounted in a heavy container on a spring suspension to minimize acoustic and mechanical disturbance.

The fractional decrease in resistance with voltage above 0.1 volt is independent of the contact resistance and whether or not the measurements are made in air or vacuum.

In curves I', II' and III' we have changed the voltage scale of the curves I, II and III to a temperature scale in accordance with the relation,

$$T = T_0 + 40V^2.$$

This relation has a theoretical basis in the Joule heating of the contacts due to the passage of current and contains the assumption of a value of Wiedemann Franz ratios characteristic of solid carbon.[5]

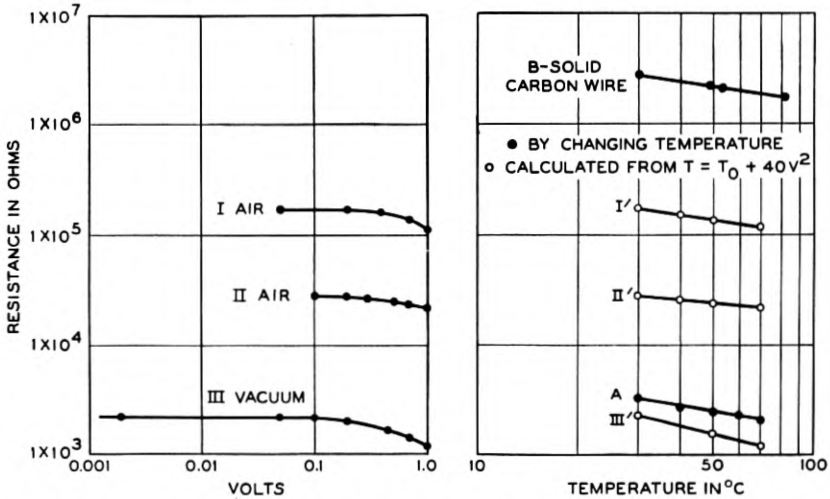

Fig. 15—Characteristics showing the effect of voltage and temperature on contact resistance.

These curves have substantially the same slope as A, which is a characteristic measured by heating a contact in the furnace, the contact voltage being sufficiently small to avoid appreciable heating of the contact due to this cause, and also with B, which was obtained with a solid carbon wire produced in a manner to simulate closely microphone carbon. We are able to conclude from measurements such as these that the nature of the conducting portions of contacts is that of solid carbon both for air and vacuum and that the departures from Ohm's law—at least up to 1 volt—are due to the Joule heating of the contacts.

From measurements similar to these in which we show that the admission of air has no effect on the temperature coefficient of resistance—although it produces a marked increase in the resistance at any particular temperature—we are also able to conclude that the presence of adsorbed air does not alter the nature of the conducting portions of the contacts but merely limits their areas.

[5] This theory, based on earlier work of Kohlrausch, was worked out in useful form independently in Bell Telephone Laboratories (unpublished work) and by R. Holm (*Zeit. Tech. Phys.*, **3**, 1922). It gives the approximate relation, const. $\frac{V^2}{K_0/\sigma_0}$, as the increase in temperature above room temperature, V being the contact voltage, and K_0/σ_0 the Wiedemann Franz ratio for the contact material.

Turning now to the effect of contact force on contact resistance: we see (Fig. 16) that large and approximately reversible resistance changes are produced as the force is varied repeatedly between fixed limits. This shows that the effect is in the main elastic, though the

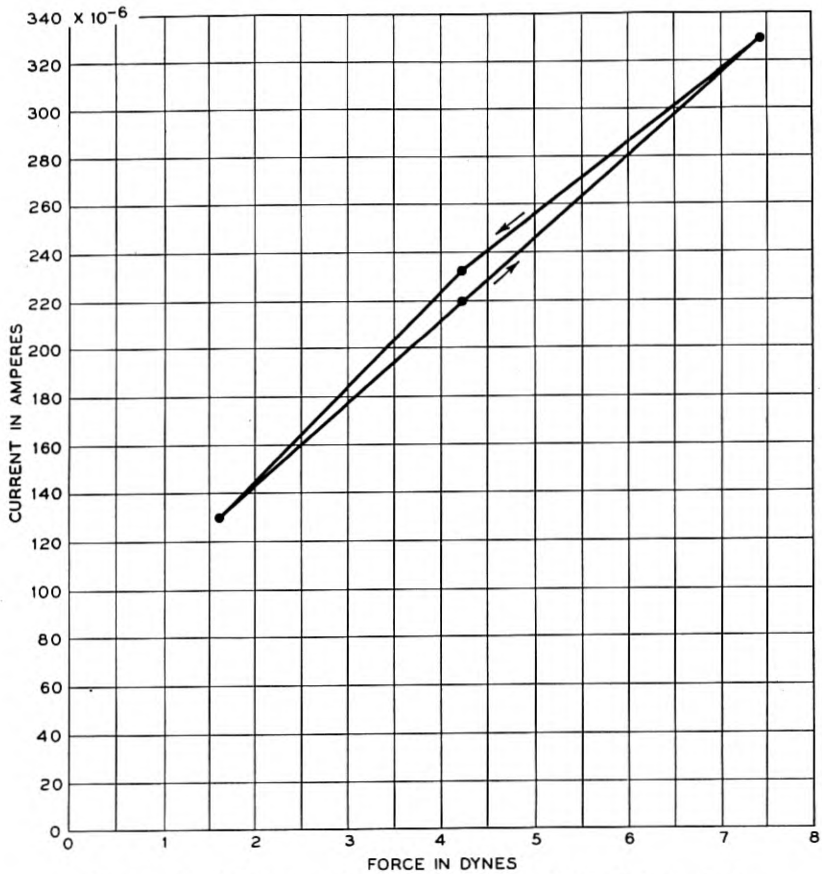

Fig. 16—Typical current-force cycle obtained with a single contact.

existence of a narrow loop indicates a small plastic or irreversible movement as a secondary effect.

We have thus established that the current is conducted through solid carbon and that the deformations are mainly elastic. These facts give strong support to the "elastic theory" of "loose contacts," i.e., the hypothesis that the change of resistance takes place because of a change in contact area under pressure. An extensive study of the resistance-force characteristics gave results which could not be simply interpreted (just as Gray had found) and, because of the possibility that unknown cohesional or frictional forces were involved, the work was extended by a study of resistance-displacement characteristics. Through a comparison of the two sets of data we were led to the conclusions that the stress-strain characteristics are not so simple as those assumed in Pedersen's or Gray's analysis and, therefore, that a study of the elastic behavior of contacts offered the most promising line of attack on the problem.

Figure 17 shows the mechanical system developed for this purpose. With it known forces can be applied to a contact element and at the same time its movement can be measured.

The contact is made between a carbon granule and a polished carbon plate, the granule being attached to the end of a rod R suspended by springs S from a fixed frame and the plate being attached to the end of a micrometer screw M_2 capable of giving to it a translational motion without rotation.

The force is applied to the granule electrostatically by means of voltage applied between the condenser plates C_2, one of which is attached to the rod R and the other to the micrometer screw. This is in principle the attracted disc electrometer of Kelvin and it is capable of applying forces up to 15 dynes without using voltages greater than 200.

The motion of the granule with respect to the carbon plate is measured electrically through the variation of capacity of the condenser C_1, of which one plate is attached to the other end of the rod R. C_1 forms part of an oscillating circuit of natural frequency n_0 (about 2000 kc.) which is coupled to a wave-meter circuit adjusted for oscillation at a frequency n_1 slightly different from n_0. Changes in the frequency arising from the changes in capacity C_1 alter the energy picked up by the wave-meter circuit and this energy, which is recorded by means of a galvanometer, serves as a measure of the change of capacity or motion of the rod R. With this arrangement it is possible to measure motions as small as 1×10^{-7} cm. and under the best conditions as small as 1×10^{-8} cm. It is necessary to have good damping, which is obtained by means of immersing the drum D in polymerized castor oil. The accessory spring S_2 is used merely for calibrating purposes.

Figure 18 shows the appearance of the apparatus as set up for measurement. The condenser is contained in the lower housing at the left, the wave-meter in the upper housing. The whole apparatus including the galvanometer is supported on a delicate spring suspension within a second large lead container, the frame of which just appears at the edge of the photograph and which is also supported by springs.

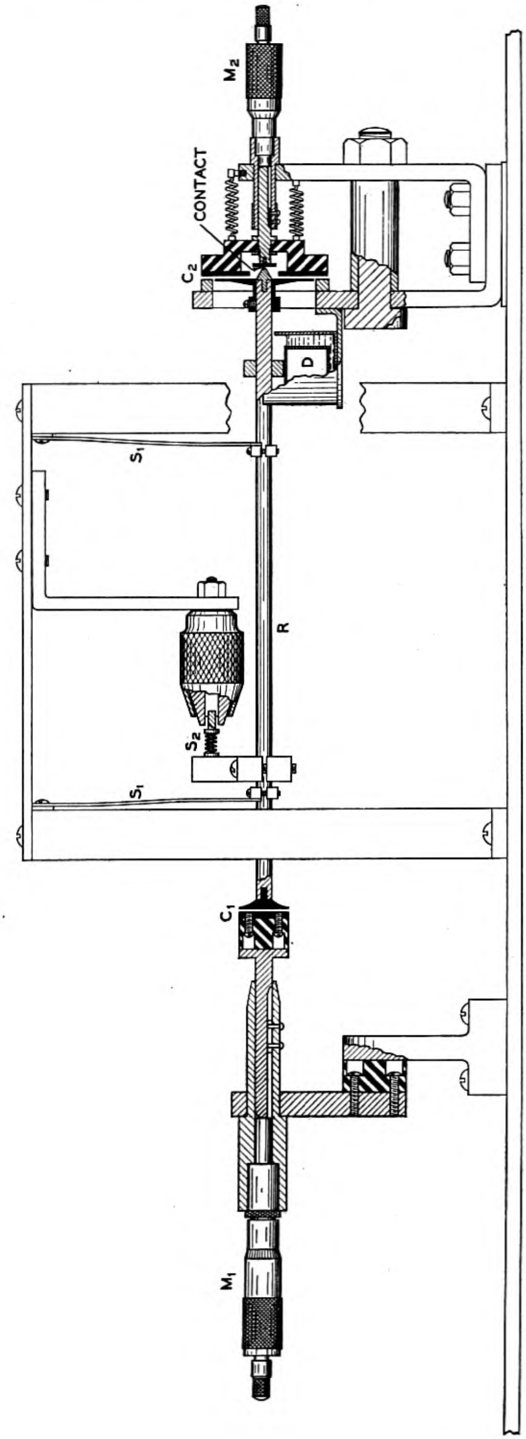

Fig. 17—Diagrammatic view of mechanical system used in the study of single contacts.

Figure 19 shows the appearance of the complete setup with the cover on the outside container. This begins to compete with cosmic ray apparatus from the point of view of the amount of lead involved, the outer container weighing about 600 lbs. Port-holes—one of

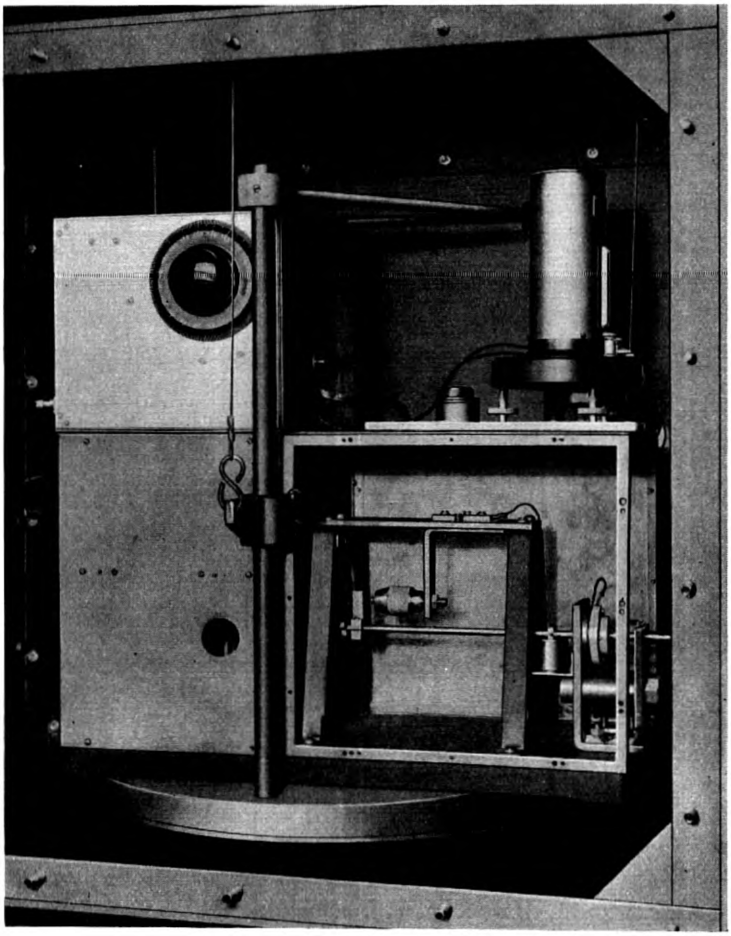

Fig. 18—Mechanical system and associated electrical apparatus as set up for single contact study.

which appears on the near end of the box—permit adjustments to be made on the apparatus within, thus eliminating the necessity for removing the large outer cover which, as you may surmise from the number of handles, requires the combined efforts of two men to remove it. All of this protection is, of course, to shield the apparatus from mechanical vibrations and acoustic disturbances.

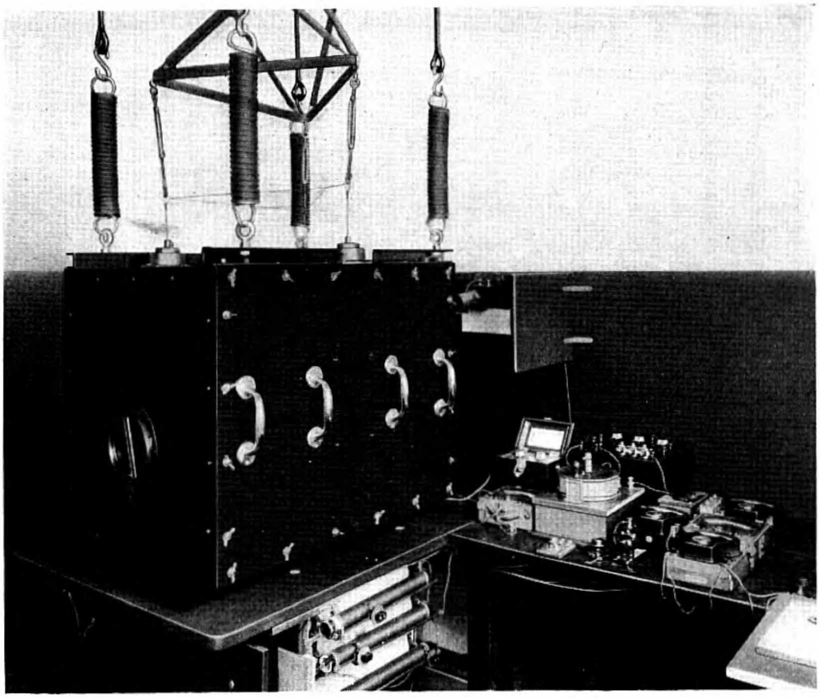

Fig. 19—Exterior view of complete experimental arrangement.

With this apparatus we investigated the variations in displacement and resistance when the forces are varied cyclically between fixed limits. Measurements on a large number of contacts are summarized in curves, Figs. 20 and 21. The cyclic characteristics, though somewhat irregular and having the form of narrow loops, approximate straight lines when the variables are plotted on logarithmic scales. Only one complete characteristic is shown in each set of curves, other typical measurements being represented by dotted straight lines joining the end points of their respective cycles. The full line in each figure represents the cycle of a typical contact, obtained by averaging, over the range in which the difference between the maximum and minimum force limits or maximum and minimum displacement limits is relatively large, in which case the slope is apparently constant.

If we let N'' and N represent the slopes of the typical force-displacement and resistance-force characteristics and if F, D and R be the contact force, the contact displacement and the contact resistance,

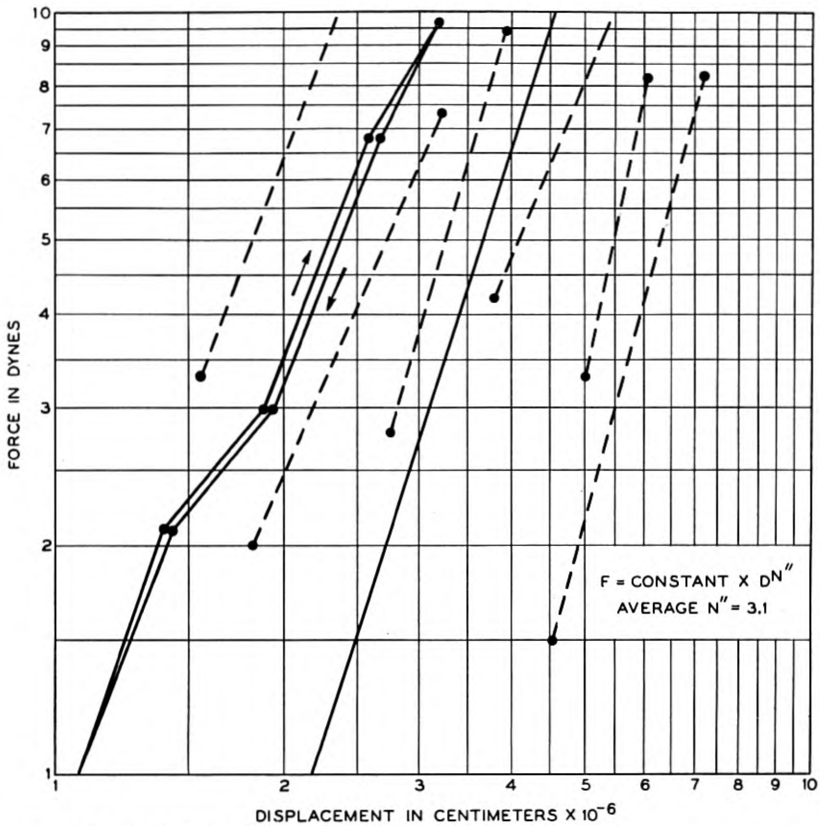

Fig. 20—Typical force-displacement characteristics of carbon granules pressed against a polished carbon plate.

respectively, we may express our results by the approximate relations:

$$F = \text{const.} \cdot D^{N''}, \tag{1}$$

$$R = \text{const.} \cdot F^{-N}. \tag{2}$$

The values N'' and N are not, however, independent of the force or displacement limits when these limits are relatively small. In Fig. 22 we have plotted values of N'' and N as functions of the difference between the maximum and minimum displacement (ΔD). We see that for relatively large values of ΔD, N'' and N approach the limiting values 3.1 and 0.47, respectively, but for smaller values of ΔD, N''

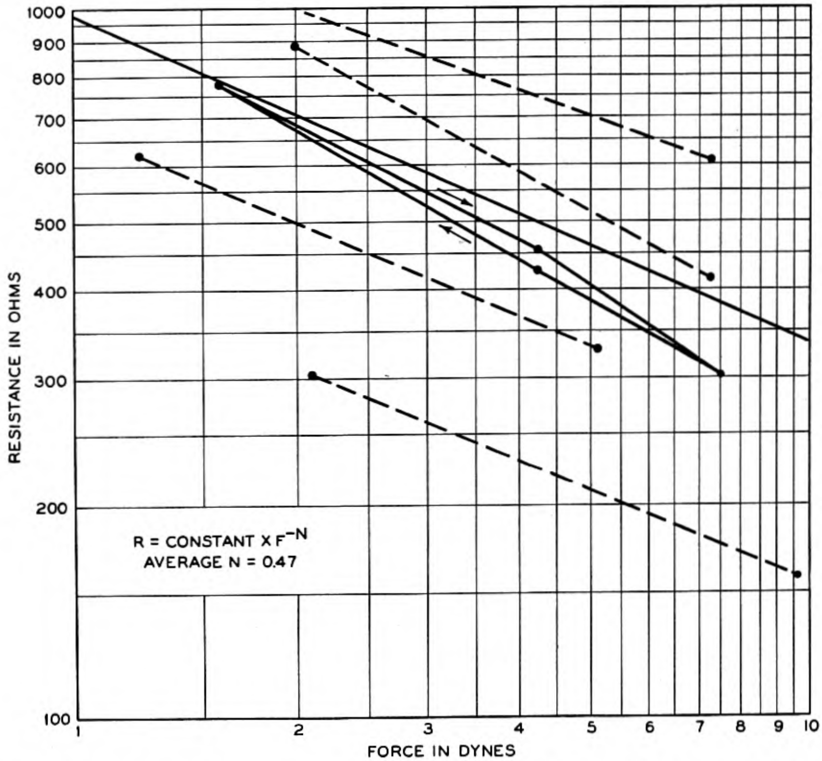

Fig. 21—Typical resistance-force characteristics of carbon granules pressed against a polished carbon plate.

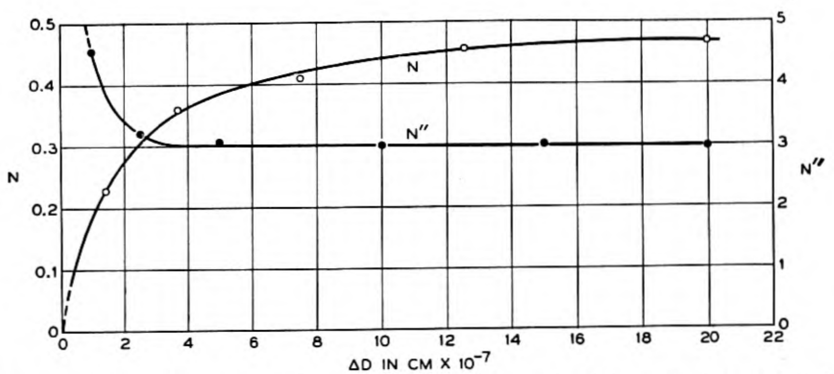

Fig. 22—Effect of the extent of contact motion (ΔD) on N and N''. (Average values.)

becomes greater than and N less than its limiting value. The limiting value of N'' is greater than that which would be obtained through the contact of hemispherical surfaces and represents a more rapid stiffening of the contact with compression.

We will first give our attention to the limiting value of N''.

A consideration of the nature of contact surfaces as revealed by the microscope furnished the clue to the interpretation of our results. A typical surface is shown in the photomicrograph (Fig. 23). Evi-

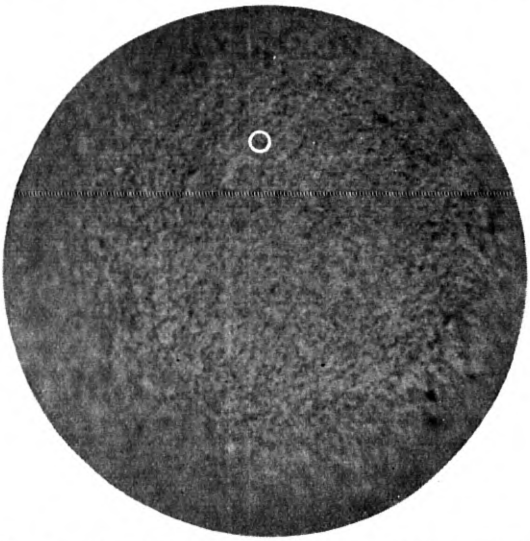

Fig. 23—Photomicrograph of the surface of a carbon granule ($\times$ 2400).

dently it is very hilly, the hills being much the same size and height. The magnification ($\times$ 2400) is such that the small white circle has a diameter of 8×10^{-5} cm. and it is clear that the circle encloses several hills.

From the theory of elasticity we may deduce that if two hemispherical hills of carbon having a radius of the order 1×10^{-5} cm. are brought together with forces of the order of 1 dyne the maximum stresses will probably not exceed the elastic limit of carbon and hence that the hills will deform elastically. The motion involved in such a deformation will be of the order of 1×10^{-6} cm. and if other hills are encountered, as is most probable with such a movement, the stresses will be shared and hence the stress per hill reduced. According to this view forces larger than one dyne can be applied without exceeding the elastic limit merely by virtue of the distribution of the hills which will come in to share the stresses. Furthermore, such a contact will stiffen up more rapidly with compressional displacement than will a contact made on a single hill. This concept of a loose contact, therefore, seemed to offer possibilities in the way of an adequate explanation of the experimental results.

At first the problem seemed too complex for mathematical analysis and a study of the elastic behavior of contact surfaces having various arrangements of little hemispherical hills was made with the aid of large scale rubber models. Quarter inch rubber balls were cut in half for this purpose and arranged on bases of suitable material and shape.

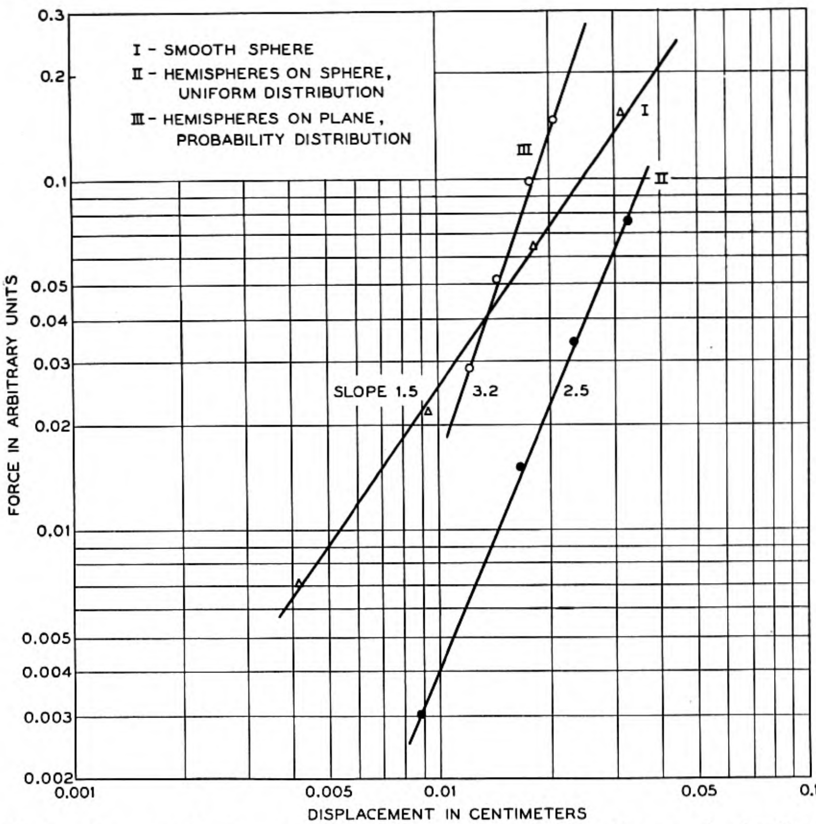

Fig. 24—Stress-strain characteristics obtained with contact surfaces made of rubber.

In Fig. 24 we have plotted the force-displacement characteristics of three different surfaces: *I*, that of a single smooth hemisphere; *II*, that of small hemispheres of equal height evenly distributed on a portion of a large 32 inch sphere made also with rubber; and *III*, that of hemispherical surfaces of random height fastened to a flat plate, about 100 hemispheres being used.

We see from the slopes of these curves that the model made with hills of random height on a flat plate behaves most like the actual contacts, the slopes of the corresponding curves being 3.2 and 3.1, respectively. This arrangement is also the one which most nearly represents the carbon surfaces as viewed under the microscope. Here the hills have various heights and the radius of the underlying base (0.015 cm.) is so much larger (1000 fold) than that of the average hill that within the region of the contact area the surface of the former may be regarded as plane.

The slope of curve I is in accord with a formula derived from the theory of elasticity by Hertz connecting the force F pressing together two elastic spheres and the movement D between the centres of the spheres:

$$F = \text{const. } D^{3/2}. \tag{3}$$

The constant includes such factors as the elastic moduli of the contact materials and the radii of the spheres and need not concern us here. The case of a sphere pressed against a flat plate, as in our experiments, is a particular case of this general equation, the constant only being affected.[6]

The slopes of curves II and III are also in accord with theory, as we shall see, when one makes the simple assumption that the elastic deformation is confined to such a small region near the contact in each hill that the underlying base is not appreciably deformed. This assumption was tested in the case of the model having the spherical distribution of hemispheres by changing the stiffness of the rubber used in the underlying sphere. No effect was produced on the stress-strain characteristic (curve II). We may therefore consider that the elastic reactions produced in each hill are independent of each other and that the base is not deformed, so that with a given distribution of hills it becomes a simple matter to calculate their combined effect over a given compressional range. We may represent the conditions essential for our calculation by the diagram, Fig. 25, in which A represents the plane surface of the smooth contact element just making contact with the highest hill of the rough contact element. Under compression, A may be considered as moving in the direction of its normal x, compressing B and, with increasing motion, coming into

[6] Formula (3) is known to hold accurately for values of D not greater than about 1 per cent. of the radius of the sphere (J. P. Andrews, *Phys. Soc. Proc.*, Vol. 42, No. 236). This condition is fulfilled in the case of curve I but D is as great as 10 per cent. of the radius in the case of a few of the hills involved in the maximum compression shown in curves II and III (Fig. 24).

contact with other hills C and compressing them according to equation (3). The position of C is conveniently defined by its distance X from the plane A.

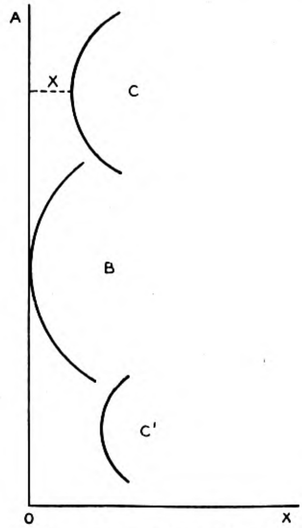

Fig. 25—Schematic representation of a rough surface used in mathematical analysis.

Any continuous distribution of hill positions, typified by C, which would be encountered through a small compressional movement, may be approximately represented by the expression,

$$N_x = \text{const. } x^n, \tag{4}$$

where N_x is defined as the number which multiplied by dx gives the number of hills coming into contact with the plane when it moves from x to $x + dx$. The exponent n is a constant which for convenience we may call the distribution constant.

For a total compression D the $N_x dx$ hills will be compressed an amount $D - x$, and hence the total force of reaction F is given by

$$F = \text{const.} \int_0^D x^n (D - x)^{3/2} dx,$$

which integrates to the form,

$$F = \text{const. } D^{n+5/2} = \text{const. } D^{N''}. \tag{5}$$

The constant here includes a summation of the individual constants of equation (3). It is clear that if the hills have different radii the constant only will be affected, so that equation (5) may be regarded as general in this respect.

For the case of uniform hills distributed on the surface of a sphere it may be shown that equal numbers of hills will be added for equal increments in x, in which case $N_x = \text{const.}$[7] From this it follows that $n = 0$ and $N'' = 2.5$ in agreement with the measured value, curve *II*.[8] For $N'' = 3.2$ as obtained with the hemispheres of random height on a plane, curve *III*, n would have the value 0.7. The corresponding distribution function N_x would approximate to that of the portion of an ordinary error curve near its maximum. A rough determination of the distribution of heights amongst the small rubber hemispheres showed in fact that they approximated closely to an error curve and that the displacement range covered that portion of the curve near the most probable height.

It would appear from this analysis that the elastic behavior of our carbon contacts under conditions of relatively large strain is adequately explained on the very simple assumption that the hills which we observe under the microscope have a random distribution of heights and behave like smooth spherical surfaces. We have, however, still to account for the hysteresis and the large values of N'' corresponding to small values of ΔD as well as the values of N (Fig. 22).

It is unlikely that the hills which we observe under the microscope are submicroscopically smooth, in which case we would expect a small plastic movement in these secondary hills arising from overstrain. We have direct evidence for this in the fact that contacts once established—even without the passage of current—require relatively small but finite forces to break them. Such junctions within the contact region could well account for hysteresis and a stiffening up of the contact in the region of small strains. Furthermore it is to be expected that they might affect the resistance behavior to a much greater extent than the elastic, and over a wider range of strain, since the junctions—though too weak to affect appreciably the contact stiffness—might well carry a relatively large proportion of current; in which case the value of N would be smaller than that calculated on the assumption of smooth spherical surfaces.

We will now derive an expression relating resistance and force for the type of contact considered in the derivation of equation (5), assuming smooth hills.

Classical theory[9] gives the following formula for the conductance

[7] This argument rests on the fact of geometry that if A is the area of contact between a sphere of radius r and a plane, $dA/dx = 2\pi r$.

[8] This agreement between theory and experiment shows that the compression of some of the hills by an amount in excess of 1 per cent. of their radii has not affected the applicability of equation (3) to our problem.

[9] Riemann Weber.

It is here assumed that the mechanical and electrical areas of contact are coincident, which according to the ideas of wave mechanics may not be the case.

$1/r$ of the contact formed by compressing, by an amount D, a single smooth conducting sphere against a flat conducting plate,

$$\frac{1}{r} = \text{const. } D^{1/2}. \tag{6}$$

It appears reasonable to assume that the hills which come into contact with compression act independently of each other as regards conduction. The conductances may therefore be added and we may write for the total conductance $(1/R)$ produced by a compression D involving many hills:

$$\frac{1}{R} = \text{const.} \int_0^D x^n (D - x)^{1/2} dx,$$

which integrates to the form,

$$\frac{1}{R} = \text{const. } D^{n+3/2},$$

which in combination with (5) gives

$$R = \text{const. } F^{\frac{2n+3}{2n+5}} = \text{const. } F^{-N}. \tag{7}$$

Using the value of n consistent with equation (5) through the measured value of N'', viz., $n = 0.6$, we get $N = 0.68$. The measured value of N (0.47) is, as we have surmised, too small though it is of the right order of magnitude.

We are, of course, investigating the factors which give rise to this discrepancy as they will play an important part in any complete theory of microphonic action, and we are extending our study to the behavior of granular aggregates in simple cells and microphone structures. We have shown that the value of N in a simple cell composed of parallel electrodes is quite consistent with our simple theory for single contacts, which therefore indicates that the behavior of an aggregate of contacts is determined by the behavior of the individual contact. Furthermore, we have shown, through static measurements on the handset instrument, that the granular aggregate within this irregularly shaped structure behaves like the aggregate in a simple cell. We are therefore confident that the behavior of the microphone will be explained in terms of the behavior of the single contact.

The behavior of the two dimensional model of the handset microphone (Fig. 26) is most convincing in this connection. Although this model was set up originally to study the distribution of stresses in this type of structure it has proved most useful in other phases of our work. Quarter inch rubber balls represent the granular particles of

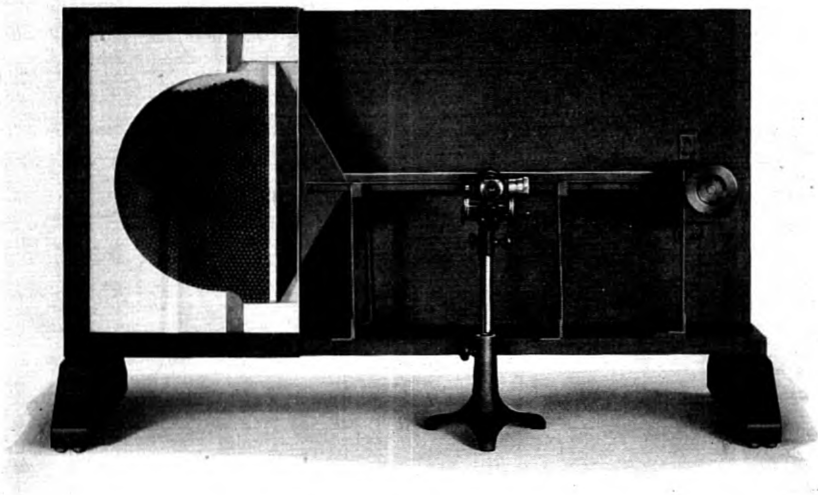

Fig. 26—Model of handset transmitter cell.

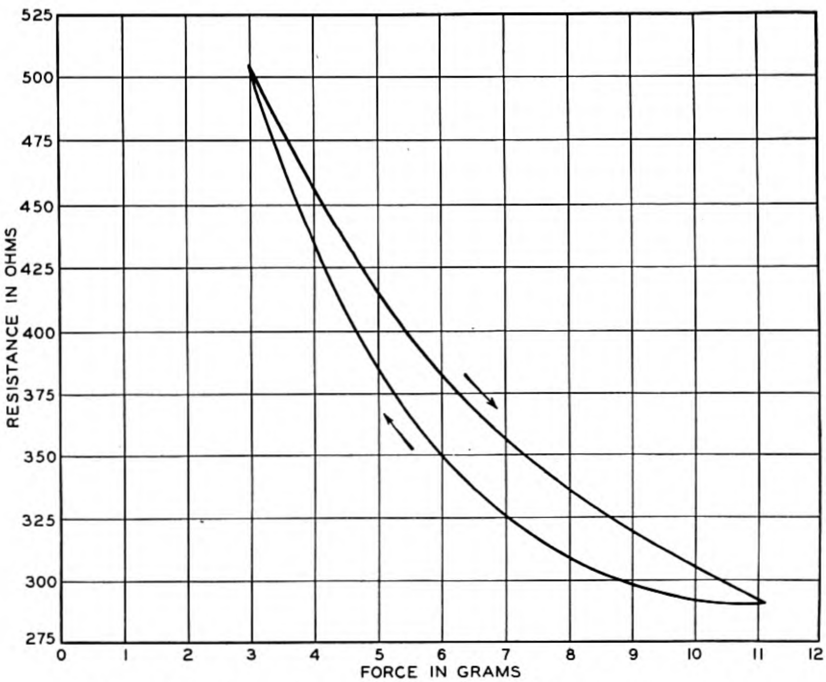

Fig. 27—Resistance-force cycle obtained with transmitter model.

the actual microphone and by coating these with a conducting layer of graphite and lacquer we are able to make them behave electrically as well as elastically in accordance with our simple theory. When placed in the model the aggregate is compressed cyclically by means of the piston which acts as a diaphragm, producing a change of resistance in the current path around the insulating barrier. The curves shown in Figs. 27 and 28 show typical resistance-force cycles, obtained with the model and the actual instrument under conditions wherein the reactive forces are mainly elastic. The similarity of these characteristics is striking. The existence of the loops indicates that the reactive forces are not entirely elastic and that the behavior is modified by friction, as in the case of single contacts.

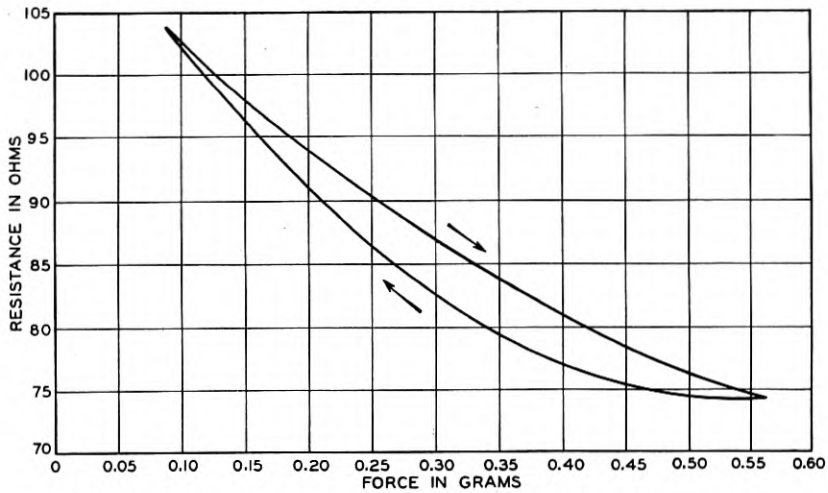

Fig. 28—Resistance-force cycle obtained with a standard transmitter.

In conclusion it seems fair to say that our experiments on "loose contacts" under conditions which are equivalent to those under which they operate in actual microphones have given a satisfactory picture of the essential nature of such contacts, and their mode of operation when strained, both from the elastic and the electrical point of view. The electrical current is carried through regions in intimate contact and changes in resistance under strain are due both to a variation in the number of microscopic hills which form the carbon surface and to area changes at the junctions of these hills arising from their elastic deformation in accordance with the well known laws of elasticity.

Open-Wire Crosstalk *

By A. G. CHAPMAN

Effect of Constructional Irregularities

IF the cross-sectional dimensions of an open-wire line were exactly the same at all points and if the transpositions were located at exactly the theoretical points, the crosstalk could be reduced by huge ratios by choosing a suitable transposition arrangement and interval between the transposition poles.

Practically, however, the crosstalk reduction is limited by unavoidable irregularities in the spacing of the wires and of the transposition poles. There is no point in reducing the type unbalances by transposition design beyond the point where the constructional irregularities control the crosstalk.

Transposition Pole Spacing Irregularities

The following discussion covers the method of estimating the crosstalk due to irregularities in the spacing of transposition poles and the derivation of rules for limiting such irregularities. With practical methods of locating transposition poles, the effect of the pole spacing irregularities may ordinarily be calculated by considering only the transverse crosstalk. Special conditions for which attention must be paid to interaction crosstalk are discussed later. The simplest case, that of transverse far-end crosstalk due to pole spacing irregularities, will be discussed first.

A transposition section is divided into segments by transposition poles which in practice vary in number from four to 128. Each segment causes an element of crosstalk current at a circuit terminal and this element is about proportional to the segment length. For far-end crosstalk between similar circuits all these crosstalk current elements would add almost directly if there were no transpositions. The function of the transpositions is to reverse the phase of half the current elements. The segments corresponding to the reversed current elements may be called the minus segments. If the other half of the

* This is the second half of a paper which was begun in the January 1934 issue of the *Technical Journal*, giving a comprehensive discussion of the fundamental principles of crosstalk between open-wire circuits and their application to the transposition design theory and technique which have been developed over a period of years.

segments are called the plus segments, the far-end crosstalk is proportional to the difference of the sum of the plus segments and the sum of the minus segments. This difference may be called the unbalanced length and the output-to-output far-end crosstalk is this length multiplied by the far-end coefficient and by the frequency.

If the sum of the plus segments equals the sum of the minus segments, the unbalanced length will be zero. The poles of a line are necessarily spaced somewhat irregularly but for a single circuit combination the unbalanced length could be made very small by carefully picking the transposition poles so as to keep the sums of the plus and minus segments about equal. This procedure is impractical, however, because many circuit combinations must be considered and because necessary line changes would prevent the maintenance of very low initial unbalanced lengths.

In practice, therefore, the segment lengths are allowed to deviate in a chance fashion from the mean segment length. The unbalanced length varies among the various circuit combinations depending on the arrangement of the transpositions which determines the order in which plus and minus segments occur. For any particular combination, the unbalanced length has a wide range of possible values and its sign is equally likely to be plus or minus.

In any transposition section, the length of any segment may deviate from the average segment length for that section. If the sum of the squares of all the deviations in each transposition section is known, the unbalanced length for a succession of transposition sections may be estimated, that is, the chance of the total unbalanced length lying in any range of values may be estimated.

Letting S_1^2 be the sum of the squares of the deviations for the first transposition section, etc., and letting R be the r.m.s. of all the possible values of the total unbalanced length in all the sections, the following approximate relation may be written:

$$R^2 = S_1^2 + S_2^2 + \cdots \text{ etc.}$$

The chance of exceeding the value R may then be computed. For example, there is about a one per cent chance that the total unbalanced length will exceed $2.6R$.

In making rules for locating transposition poles the first step is to determine a value for R. For example, if consideration of tolerable crosstalk coupling indicated that there should not be more than one per cent chance that the total unbalanced length in a 100-mile line would exceed one mile, then R, the r.m.s. of all possible values of the total unbalanced length, should not exceed 1/2.6 miles. Since R is calculated from the values of S for the individual transposition sections, a given permissible value of R may be obtained with various sets of values of S. It seems reasonable to determine individual values of S on the principle that a transposition section of length L_s should have the same probability of exceeding a given unbalanced length as any other section of the same length and that a section of length $2L_s$ should have the same probability as two sections of length L_s, etc. On this basis, the value of S^2 for any transposition section should be proportional to the section length L_s. This leads to the rule used in practice that for any transposition section S^2 should not exceed kL_s. If L_s and S are expressed in feet, a value of three for k is found suitable for practical use. The choice of a value for k will depend, of course, upon the cost of locating and maintaining transposition poles with various degrees of accuracy and upon the effect on the crosstalk of varying the value of k.

The above rule permits a large deviation at one point in a transposition section if it is compensated by small deviations in the rest of the segments. For example, with 128 segments and a mean segment length of 260 feet, one long segment of 575 feet is permissible if the rest of the segments are 258 feet. The expression for the total unbalanced length in a succession of transposition sections assumed that the deviations varied from segment to segment in a truly random manner. The above example involves an unusual arrangement of the deviations. When there are a number of transposition sections in a line, such unusual arrangements of deviations in various sections do not have much effect on the probability that the total unbalanced length will exceed a given value.

The computation of near-end crosstalk due to pole spacing irregularities is a more complicated problem since the crosstalk elements resulting from the various segments vary in their magnitudes and phase relations because the various segments involve different propagation distances. It may be concluded, however, that the r.m.s. value of the total unbalanced length in all the sections may be expressed as follows:

$$R^2 = S_1^2 + S_2^2A_1^2 + S_3^2A_2^2 + \cdots.$$

This differs from the expression for far-end crosstalk in that the values of S^2 for the second and succeeding transposition sections are multiplied by attenuation factors. The attenuation factor A_1^2 corresponds to propagation through the first section to the second section and back again. The other attenuation factors are similarly defined. The above expression neglects attenuation within any particular transposition section since this is ordinarily small. It also assumes that the rule for locating transposition poles, that is, that S^2 should not exceed kL_2, is applied for lengths having only negligible attenuation.

In making estimates of R in connection with transposition design work, it is assumed that all the segments are nominally the same length, D, and that r is the r.m.s. value of the deviations of the segments. Since r^2 equals S^2 divided by the number of segments in length L_s, r^2 should not exceed kD. R^2 may be expressed approximately in terms of r^2 as follows:

$$R^2 = r^2 \frac{1 - \epsilon^{-4\alpha L}}{1 - \epsilon^{-4\alpha D}},$$

where R and r are expressed in the same units, L is the length of the line in miles, α is the attenuation constant per mile, and D is the segment length in miles. If the line loss is 6 db or more the expression is nearly equal to:

$$R^2 = \frac{r^2}{.46Da},$$

where a is the line loss in db per mile and D is the segment length in miles. This assumes $4\alpha D$ is small compared to unity which is usually the case.

The chance that the total unbalanced length will exceed about $2.1R$ is estimated at 1 per cent.

For far-end crosstalk (output-to-output) the same assumption as to nominal segment length leads to the expression:

$$\frac{R}{r} = \sqrt{\frac{L}{D}}.$$

The general expressions given for R^2 suggest that a very long segment might be permitted at some point in the line if the deviations of the segments were properly restricted in other parts of the line. The expressions given for far-end and near-end values of R^2 were

$$R^2 = S_1^2 + S_2^2 + S_3^2 + \cdots,$$

$$R^2 = S_1^2 + S_2^2 A_1^2 + S_3^2 A_2^2 + \cdots.$$

If a very long segment at some point, such as a river crossing, were permitted, this would increase the sum of the squares of the deviations for some transposition section. For example S_3 might be abnormally large. R^2 could be kept at some assigned value by limiting S_1^2, S_2^2, etc. This procedure is not considered good practice because of the difficulty of maintaining some parts of the line with very small deviations of the segments from their nominal lengths.

A very long segment has another effect on near-end crosstalk not indicated by the above discussion. If there were no deviations in any of the segments, the near-end crosstalk would be the vector sum of a number of current elements of various magnitudes and phase angles and the sum would be small due to a proper choice of these magnitudes and angles in designing the transpositions. If a segment deviates from its normal length, the magnitude of the crosstalk due to the segment changes and the phase angle also changes. The phase angles of the crosstalk values due to succeeding segments are also changed since they must be propagated through the segment in question. For ordinary deviations in segment lengths these effects on the phase angles may be neglected.

Since transverse crosstalk is independent of transpositions occurring in both circuits at the same point, it would appear from the above discussion that the location of such transpositions need not be accurate. This is not ordinarily a question of practical importance. If some circuit combinations have both circuits transposed at a certain transposition pole there will usually be other combinations which have relative transpositions at this pole. The transposition pole is of importance, therefore, in connection with the latter combinations and the same accuracy of location is required for all transposition poles. A question of practical importance, however, is whether the above rules for locating transposition poles properly limit the interaction crosstalk. This is affected by transpositions in both circuits at the same pole as well as by relative transpositions. In the following discussion of this matter it is concluded that the effect of transposition pole spacing irregularities on interaction crosstalk may be ignored at frequencies now used for carrier operation.

The effect of deviations in segment length on interaction crosstalk is indicated by Fig. 18. This figure indicates a short part of a parallel between two long circuits a and b. A representative tertiary circuit c is also shown. The transposition arrangements are like those of Fig. 9B. In connection with the latter figure it was shown that the interaction crosstalk would be very small if all segments had the same length d. On Fig. 18, D is used to indicate the normal segment length and the deviation of two segments from D is indicated by d. Since the length AC equals the length CF, these deviations have no effect on the transverse crosstalk which is controlled by the transposition at C. The deviations affect the interaction crosstalk between the length CF and length AC.

The circuit a has near-end crosstalk coupling with circuit c in the length CF. This effect is normally practically suppressed by the transposition in a at E. Due to the deviation d of segment CE, the near-end crosstalk between a and c in length CF will not be suppressed but will be proportional to d. There will likewise be near-end crosstalk between c and b in the length AC proportional to d. The two deviations, therefore, introduce interaction crosstalk practically proportional to d^2.

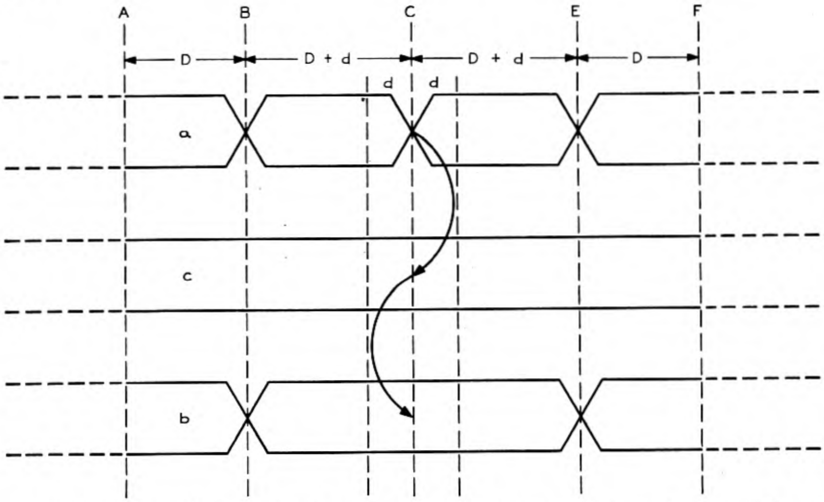

Fig. 18—The effect of deviations in segment length on interaction crosstalk.

Since there will be small deviations in numerous other segments of circuit b, the deviation d in circuit a will introduce numerous other interaction crosstalk paths similar to that discussed above. The r.m.s. value of the total interaction crosstalk caused by deviations in segment lengths may be roughly estimated as follows:

$$\frac{2FK\gamma r^2\sqrt{\frac{L}{D}}}{\sqrt{.46aD}} = \frac{.1FK^2r^2\sqrt{\frac{L}{D}}}{\sqrt{aD}},$$

where r is the r.m.s. deviation, L is the line length, D is the nominal segment length and a is the line loss in db per mile, all distances being expressed in miles. The above expression varies about as the 1.75 power of frequency and as the square of r. The corresponding expression for transverse crosstalk, i.e., $FKr\sqrt{\frac{L}{D}}$ varies as the first power of frequency and of r. It follows that, if the rules for accuracy of transposition pole spacing are relaxed or the maximum frequency is raised, the effect of pole spacing on interaction crosstalk increases more rapidly than the effect on the transverse crosstalk.

For the range of frequencies and accuracy of pole spacing used in practice, it has been found that the effect of pole spacing irregularity on interaction crosstalk is not controlling. This is indicated by Fig. 19

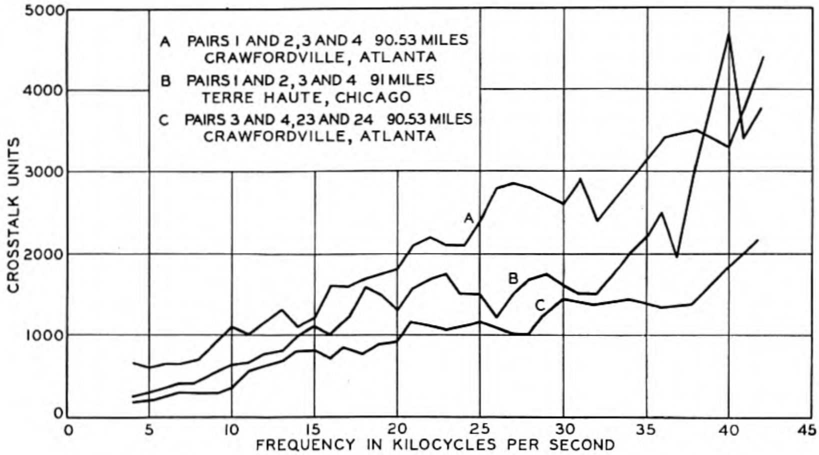

Fig. 19—Far-end crosstalk caused by pole spacing irregularities.

which shows some measurements of output-to-output far-end crosstalk between long circuits having transposition arrangements designed to make the crosstalk due to type unbalance small compared to that due to irregularities. The curves are about linear with frequency as would be predicted if the effect of the pole spacing irregularities (and wire spacing irregularities) on the interaction crosstalk is neglected. For these particular curves, a knowledge of the pole spacing indicated that pole spacing rather than wire spacing irregularities were controlling in causing crosstalk.

The above discussion assumes that a transposition section is divided by the transposition poles into segments all of the same nominal length. It is sometimes economical to use segments of different nominal lengths in the same transposition section. If the variation among the segment lengths is consistent rather than accidental it may be allowed for in the design of the transpositions.

In practice, segments of different lengths are used in the same transposition section when it is desired to adapt for multi-channel carrier frequency operation a few pairs on a line already having many pairs transposed for voice-frequency operation. Such lines often have existing transposition poles nominally spaced ten spans apart while for the pairs retransposed for carrier operation it is necessary to space the transposition poles about two spans apart. In such cases the cost of the carrier channels is appreciably increased if uniform spacing between the new transposition poles is used. The transpositions in the pairs retransposed for carrier operation must be coordinated with the transpositions in the other circuits and it is necessary, therefore, either to divide the ten spans into four approximately equal parts with consequent expense in setting new poles at the quarter points or to retranspose all the circuits on the line.

To avoid either of these expensive procedures, the new transposition poles are nominally located in the manner indicated by Fig. 20. This

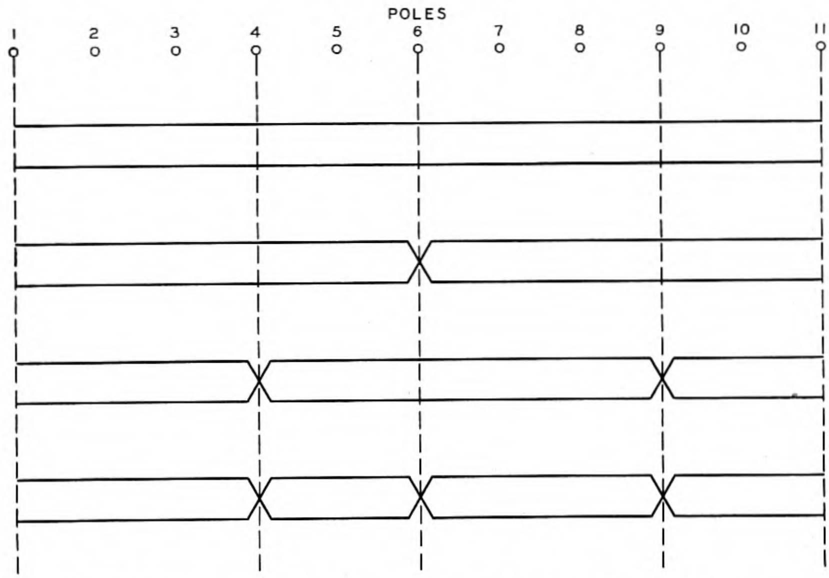

Fig. 20—Location of extra transpositions in a ten-span segment of line.

figure shows ten-pole spans subdivided into four parts in order to create three additional transposition poles. The figure indicates the location of the new transposition poles and the possible methods of transposing at these new poles. For some of the circuit combinations the crosstalk within the ten-span interval is considerably greater than if the four segments were equal in length. In each other ten-span interval the crosstalk is likewise increased by a similar inequality in segment length. Since all ten-span intervals are nominally alike, considerable crosstalk reduction may be obtained by properly designed transpositions located at the junctions of these intervals.

The use of segments of different lengths inherently decreases the effectiveness of the transpositions in reducing crosstalk and adds to the complexity of the transposition design problem. Uniform segments are therefore used except in special circumstances.

Wire Spacing Irregularities

In the past there has been a tendency to permit wire spacing irregularities in order to reduce the cost of construction and maintenance. For example, "H fixture" crossarms formerly had special wire spacing to permit the two poles to pass between pairs of wires and thus reduce the length of the arms. Another example is that of resetting a pole with a rotted base and reducing the spacing between crossarms to get clearance between wires and ground. The development of repeatered circuits and carrier current operation has increased the seriousness of the crosstalk resulting from such irregularities and made such practices generally undesirable.

There are, of course, unavoidable irregularities in wire spacing due to variations in dimensions of crossarms, insulators and pole line hardware and warping of crossarms. Corners and hills are other causes since the crossarms at a corner and the poles on a hill are not at right-angles to the direction of the wires. The most important unavoidable spacing irregularity is, however, due to variations in wire sag. Of recent years, limits have been set on wire sag deviations to insure that this effect is properly limited during construction. The main criterion adopted has been the difference in sag of the two wires of a pair. This difference is a rough measure of the crosstalk increment due to variations of the sag from normal. The crosstalk between two pairs in a given span will be abnormal if the two pairs have different sags even if there is no difference in sag for the two wires of a pair. The crosstalk is usually more nearly normal, however, than in the case of two pairs having the same average sag but different sags for the two wires of a pair. As far as practicable, all pairs are sagged alike in a given span.

The crosstalk between two pairs due to sag differences is computed much like that due to pole spacing irregularities. The change in crosstalk due to a known pole spacing deviation may, however, be computed from the crosstalk coefficient while the change in crosstalk due to a sag deviation is not related to the crosstalk coefficient in any simple way. Two methods have been used to obtain constants for calculation.

With the first method, crosstalk measurements were made on a long line (about 100 miles) having small pole spacing and type unbalance crosstalk. The r.m.s. of a number of crosstalk measurements was determined for each particular type of pair combination, for example, for horizontally adjacent pairs. The r.m.s. of the sag differences in a representative number of spans was also determined for the two pairs of each type of combination. The two r.m.s. values for any particular type of pair combination were called R and r. The ratio of R to r gave a constant k for estimating R from a known value of r and for L_0, the particular length of line tested. For other line lengths, R is estimated from the expression $R = kr\sqrt{\frac{L}{L_0}}$. Having computed R, the chance of the crosstalk for any pair combination in a long line lying in a given range may be estimated by probability methods.

The second method of studying sag differences is more precise although much more laborious. The change in crosstalk due to introducing sag differences in but two spans is determined. The poles are specially guyed to make it possible to adjust all the wires in these spans to have practically the same sag. Turnbuckles are installed at the ends of the two-span interval for this purpose. At the center pole the wires were supported so as to slip readily and equalize the sag in the two spans.

The phase and magnitude of the crosstalk is first measured for all pair combinations with all wires at normal sag. The wires are terminated in the same way as in the measurements of crosstalk coefficients. From sag measurements on actual lines, a set of unequal sag values for *all* the wires is then selected by probability methods and the crosstalk remeasured. The vector difference between the values of crosstalk before and after introducing unequal sags is then determined. This process is repeated a large number of times in order to cover the range of sag conditions encountered in practice. An r.m.s. value of the change in crosstalk due to sag difference is then determined for each pair combination and related to the r.m.s. sag difference per pair. This permits the probable crosstalk in a long line to be estimated and the importance of sag difference crosstalk to be determined. The two methods of study were found to be in general agreement. The second method has been extensively used to study proposed new wire configurations.

Drop Bracket Transpositions

An ideal transposition would cross the two sides of a circuit in an infinitesimally small distance, there being no displacement of the wires from their normal positions on either side of the transposition. The point-type transposition indicated by Fig. 21 is close enough to the ideal for practical purposes. Its deviation from the ideal requires little consideration in transposition design. To avoid cutting the wires, one wire is raised about 3/4 inch and the other lowered this amount at the transposition point. The drop bracket transposition

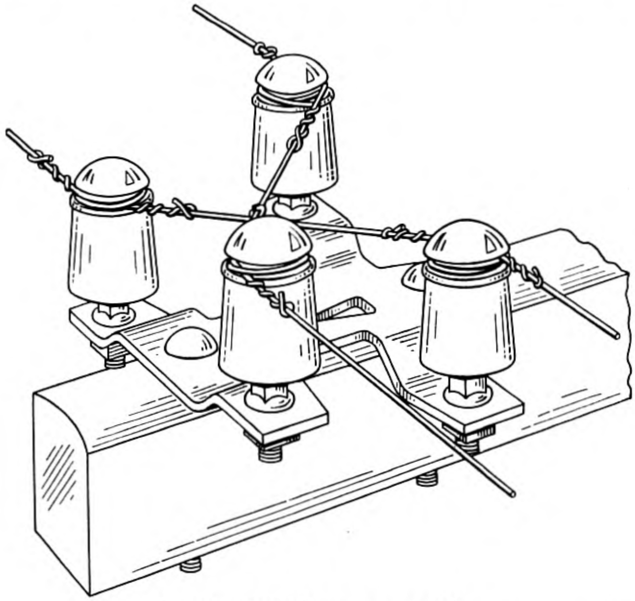

Fig. 21—Point-type transposition.

illustrated by Fig. 22 is considerably cheaper but the displacement of the wires is much greater. The effect of this displacement is important and must be especially considered in transposition design.

If all the spans adjacent to a drop bracket were of the same length

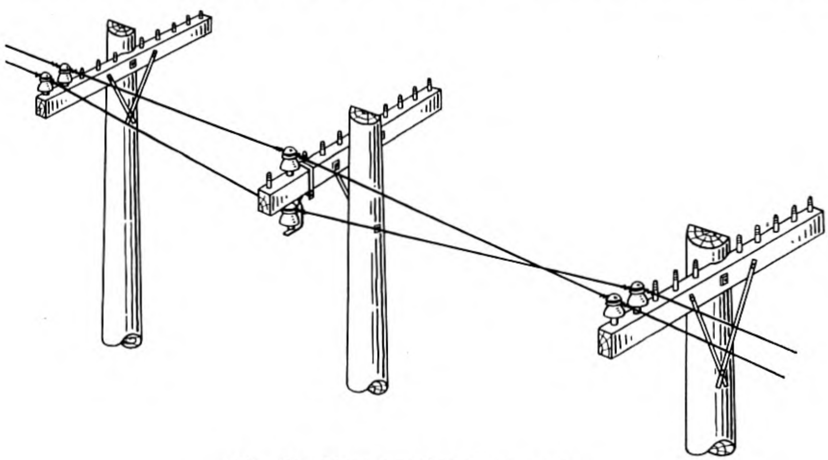

Fig. 22—Drop-bracket transposition.

and all wires could be kept under the same tension, the effect of drop brackets on crosstalk would be consistent and could, theoretically, be made negligible by a suitable transposition design.

There is, however, an accidental crosstalk effect. This effect is partly due to the fact that it is more difficult to avoid deviations from normal sag in the spans adjacent to drop brackets than in normal spans. The main effect, however, is thought to be due to inequalities in the lengths of the spans adjacent to drop brackets.

The crosstalk in such a span is very nearly proportional to the length of the span times a constant or "equivalent crosstalk coefficient." The usual crosstalk coefficient can not be used because the wires are not parallel.

Fig. 23-A indicates two long circuits, one circuit being transposed on drop brackets at the first and third quarter points of the short length D. The lengths of the spans adjacent to the drop bracket transpositions are indicated by d_1 to d_4. The equivalent far-end crosstalk coefficient for the span preceding a transposition bracket is F_1 and that for the span following the bracket is F_2. (F_1 and F_2 are usually quite different.) The total far-end crosstalk (output-to-output) due to the four spans is (very nearly):

$$K(F_1d_1 - F_2d_2 - F_1d_3 + F_2d_4),$$

where K is the frequency in kilocycles.

If the four spans were equal the crosstalk would be zero (very nearly). The actual value of the crosstalk is a matter of chance since the deviations of the four spans from the normal length are a matter of chance. These deviations cause a chance increase in the near-end crosstalk as well as in the far-end crosstalk.

This effect has been studied experimentally by using transposition designs which suppressed the consistent effect. The pole spacing effect was minimized by using very accurate spacing. The wire sag effect was allowed for by comparing similar pair combinations transposed alike except that dead-ended point transpositions were compared with drop bracket transposition. Due to the great number of transpositions necessary at carrier frequencies it was found that the accidental drop bracket effect was important at these frequencies. In recent years, point-type transpositions have been extensively used on lines transposed for long-haul carrier systems.

When, for economic reasons, a transposition system is designed for use with drop bracket transpositions, the consistent crosstalk effect must be considered in the transposition design. The equivalent crosstalk per mile for a span adjacent to a drop bracket must be determined for each pair combination. Approximate methods of computation have been worked out for doing this and checked against measurements. The computations are involved in connection with far-end crosstalk since the "tertiary effect" is controlling. Since the

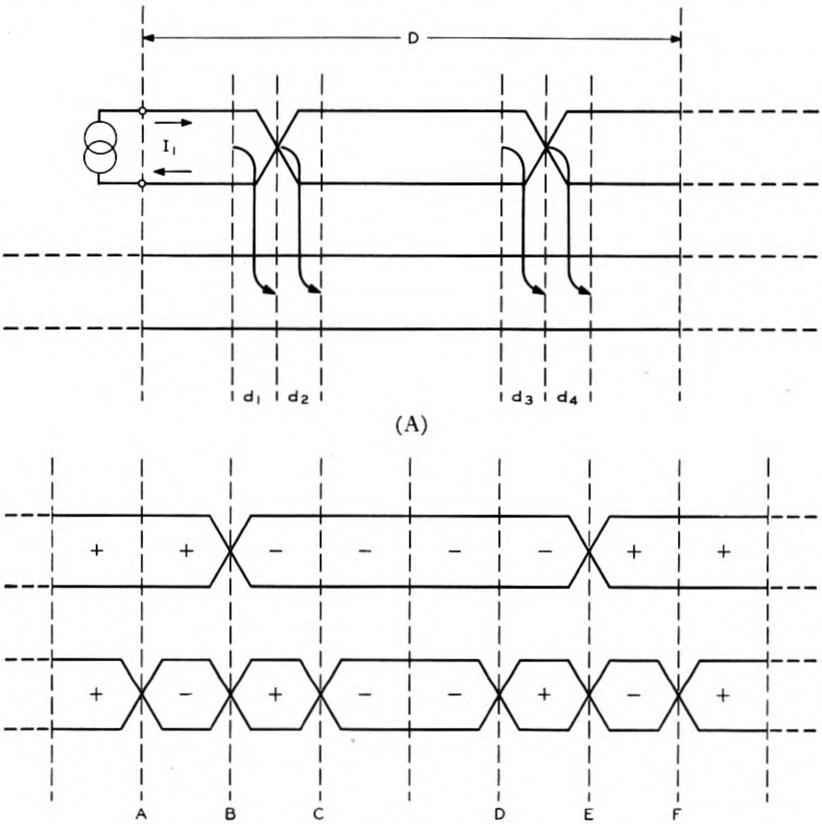

Fig. 23—Effect of drop brackets on crosstalk.

summation of crosstalk due to drop brackets is a consistent effect, "drop bracket type unbalances" can be worked out and used in transposition design. This matter is so complicated, however, that the practical method of design is to first practically ignore the drop bracket effect and then check the design to determine whether this effect has been properly suppressed.

Certain rules are adopted, however, to ensure that the transposition arrangements are properly chosen to avoid the larger drop bracket effects. Fig. 23-B indicates an arrangement of transpositions for two pairs in a short length of line which, with point transpositions, would have very low crosstalk. At points B and E both circuits are transposed alike. With point transpositions the near-end crosstalk in the two spans adjacent to one of these pairs of transpositions would be $NK2d$, where d is the span length, N the near-end crosstalk coefficient and K the frequency in kilocycles. For drop bracket transposition the crosstalk would be $K(N_1 + N_2)d$ or a change of $K(N_1 + N_2 - 2N)d$.

The transpositions are so arranged that the crosstalk in the two spans at B tends to add to that in the two spans at E. With drop brackets at B and E the major crosstalk in this length of line would be twice the above change since the crosstalk with point transpositions is very small.

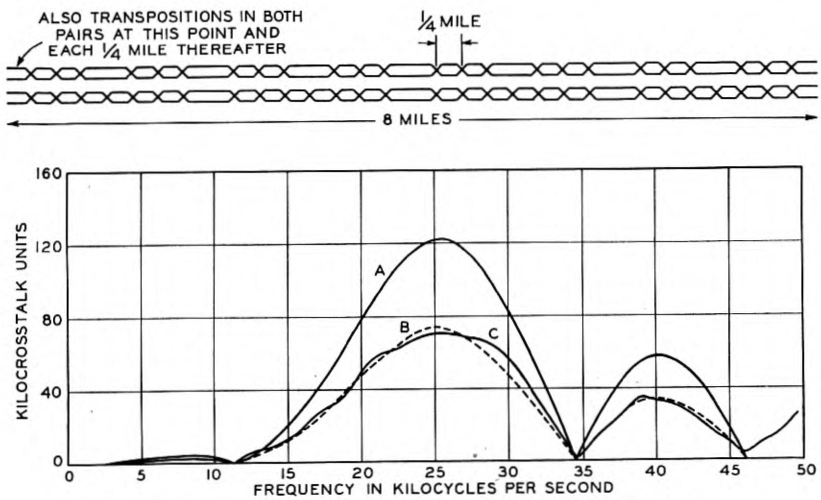

Fig. 24—Near-end crosstalk with and without drop brackets.

If the arrangement of Fig. 23-B is reiterated in a long line, the total increase in the crosstalk due to drop brackets at such points as B and E may be marked. It may be noted that the crosstalk in the two spans at A tends to cancel the crosstalk in the two spans at C and likewise there is cancellation at D and F. Drop brackets may, therefore, be used at points A, C, D and F without a consistent increase in crosstalk. Arrangements like those at B and E of Fig. 23B should be avoided in transposition design involving drop brackets.

The change in the crosstalk due to drop brackets is not necessarily an increase. Fig. 24 shows an arrangement of transpositions in an eight-mile line and three crosstalk frequency curves. Curve A shows the calculated near-end crosstalk for ideal point transpositions. Curves B and C show the calculated and observed near-end crosstalk for drop bracket transpositions. The curves show that the drop bracket effect can be calculated quite accurately and that it may reduce the total crosstalk. In the general case, it is impractical to take much advantage of this reduction effect because a marked reduction for one combination of circuits is likely to result in an increase for some other combination and because a reduction of crosstalk in one part of the line may increase the vector sum of crosstalk elements from all parts of the line.

Wire Configurations

The crosstalk coefficients for the various pair combinations may be altered by changing the configuration of the wires. Therefore, the crosstalk for a given transposition design and a given accuracy of transposition pole spacing irregularity may also be altered. The crosstalk due to sag differences also depends on the wire configuration. It is important, therefore, to choose a configuration most desirable from the crosstalk standpoint. Such an optimum configuration requires the fewest transpositions and least accuracy of pole spacing for a given maximum frequency and given permissible values of crosstalk coupling.

Various "non-inductive" arrangements of wire configurations have been suggested and tested. Such arrangements may appear to have possibilities but their study to date has indicated that they are impracticable for more than a few pairs on a line.

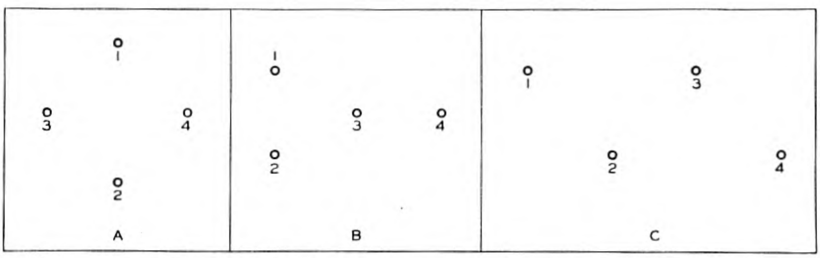

Fig. 25—"Non-inductive" arrangements for two pairs of wires.

Fig. 25 illustrates several suggested arrangements for two pairs. Arrangement A is often called a square phantom. If pair 1–2 is the disturber and there are equal and opposite currents in wires 1 and 2 there will be no voltages induced in either wire 3 or wire 4 because either of these wires is equally distant from wires 1 and 2. Since wires 1 and 2 are not equally distant from the ground, the currents in these wires may be not quite equal and opposite. As a result, voltages will be induced in wires 3 and 4 but these will be equal and there will be no crosstalk current in pair 3–4. By the reciprocal theorem the crosstalk between the two pairs will also be zero when pair 3–4 is the disturber.

Arrangement B is nearly non-inductive. In this case if pair 1–2 is the disturber and the currents in the two wires are not quite equal and opposite due to the presence of the ground, unequal voltages will be induced in wires 3 and 4 and there will be a crosstalk current in this pair. This effect could be minimized by transposing both pairs at the same points. They would not require relative transpositions since equal and opposite currents in pair 1–2 will induce no voltage in either wire 3 or wire 4.

With pair 3–4 as the disturber, equal and opposite currents will result in equal voltages induced in wires 1 and 2. These voltages cause a phantom current in phantom 1–2/3–4. This phantom current will divide between wires 3 and 4 but can not induce unequal voltages in wires 1 and 2 because 1 and 2 are equally distant from either 3 or 4. The crosstalk coefficient is, therefore, zero both for the direct effect and for the indirect effect of the phantom. However, the indirect effect of the ground or other conductors is not zero and may require transpositions.

Arrangement C is non-inductive for direct crosstalk. It is not non-inductive in regard to the indirect effect of the phantom 1–2/3–4. Equal and opposite currents in pair 1–2 induce equal voltages in wires 3 and 4. The resulting equal phantom currents in wires 1 and 2 of phantom 1–2/3–4 will induce unequal voltages in pair 3–4.

When there are many pairs on a line it is not possible to make all combinations strictly non-inductive even for direct crosstalk. With perfect wire spacing the larger values of direct crosstalk per mile could be greatly reduced, however, and appreciable reductions could be obtained in the indirect effect which is usually controlling in far-end crosstalk.

Wire sag deviations must be considered, however. If a given number of "non-inductive" pairs are placed in the pole head area normally occupied by the same number of pairs with conventional configuration, the crosstalk due to sag deviations is likely to be more serious with the "non-inductive" pairs than with conventional pairs. For the same pole head area, the number of transpositions and, therefore, the "pole spacing" crosstalk could be reduced if non-inductive arrangements were used. The tests to date indicate, however, that the total crosstalk would not be reduced because of increased "sag difference" crosstalk.

The mechanical problem of supporting the wires of the "non-inductive" arrangements is considerable if serious increases in crossarm and hardware costs are not to be incurred. This objection seems at present to override the possible advantages of (1) fewer transpositions for a given pole head area and crosstalk result, or (2) fewer transpositions and lower crosstalk with a greater pole head area.

Another possibility is the use of non-parallel wires. It is possible to arrange two pairs of wires in such a way that they have a certain direct crosstalk per mile at one end of a span and the value at the other end of the span is about equal and opposite. The net direct crosstalk per mile integrated over the span is zero or small. An example of this is the barreled square formerly used abroad. Fig. 26 illustrates this arrange-

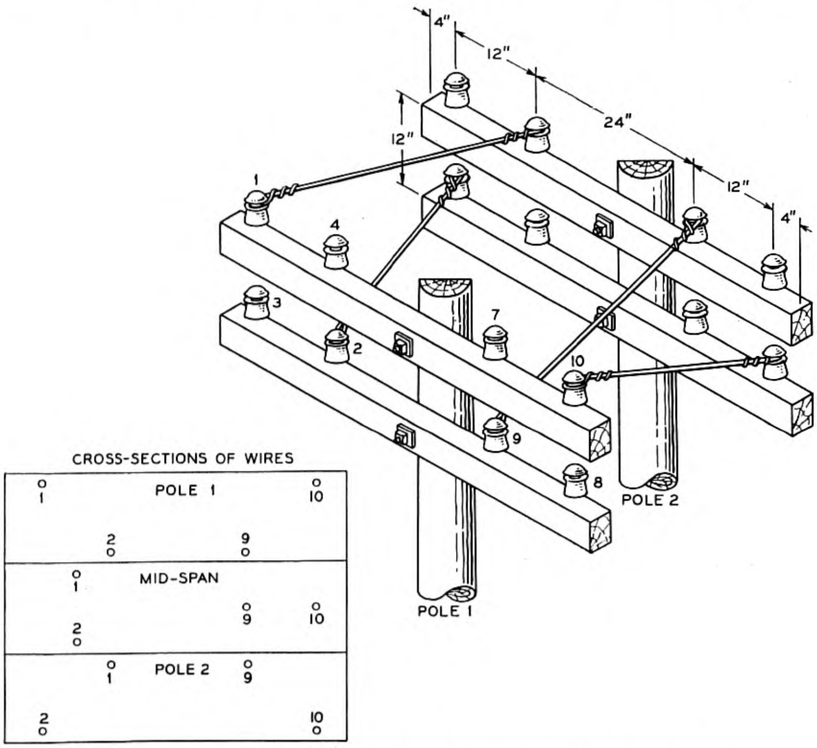

Fig. 26—Two pairs of wires in different barreled squares.

ment. The wires are arranged in groups of four, each four being arranged on the corners of a square. The two wires of a pair are on diagonally opposite corners of a square. Each pair is given a quarter turn in each span. For simplicity only two pairs in different four-wire groups and one span are shown. The two pairs shown are nearly "non-inductive" for direct crosstalk in this span.

Consideration has been given to applying this principle to a number of pairs in order to reduce the crosstalk coefficients. Since all the crosstalk coefficients could not be made very small, transpositions would be needed. The experience to date indicates that this method does not look attractive because it is not very effective in reducing the indirect crosstalk, the mechanics of transposing are difficult, the variations in sag are likely to be abnormal and the system is complicated.

There remains the simple method of improving the configuration of the wires in a given pole head area by reducing the spacing between the wires of a pair and increasing the spacing between wires of different pairs.

The crosstalk per mile between pairs is evidently reduced by this procedure since the two wires of a pair are approaching the ideal of being equally distant from every other conductor. The "sag difference crosstalk" is also reduced and higher frequencies may be used for a given crosstalk result. Fig. 27-A and Fig. 27-B indicate a 20-wire line with the wire spacing used in the past and also the configuration commonly used today on lines where heavy carrier development is involved. The spacing between the two wires of a pair has been reduced from 12 inches to 8 inches and the spacing between pairs correspondingly increased.

It was not possible to reduce the spacing of the pole pairs and for this reason they are unsuited for the higher carrier frequencies and it is sometimes uneconomic to string them. For such cases the crossarm indicated by Fig. 27-C may be considered. The 8-inch spacing of pairs is retained but the distance between pairs is further increased. With this last crossarm, phantom circuits are not superposed on the 8-inch pairs since their use results in greater crosstalk between the pairs and restricts the possibilities of multi-channel carrier operation.

The crossarm with 8-inch pairs and pole pairs may be used on lines where multi-channel carrier operation is not employed. In such cases, the 8-inch pairs may be phantomed. Since the average spacing between the side circuits of such a phantom is not reduced by the 8-inch spacing, the crosstalk between the phantom circuits is about the same as with the 12-inch pairs. The crosstalk from a side circuit into a phantom is somewhat reduced because of the reduced spacing of the pairs. For a given pole head area it does not appear practicable to devise a configuration which will result in marked reductions in the susceptibility of both phantoms and side circuits to crosstalk and

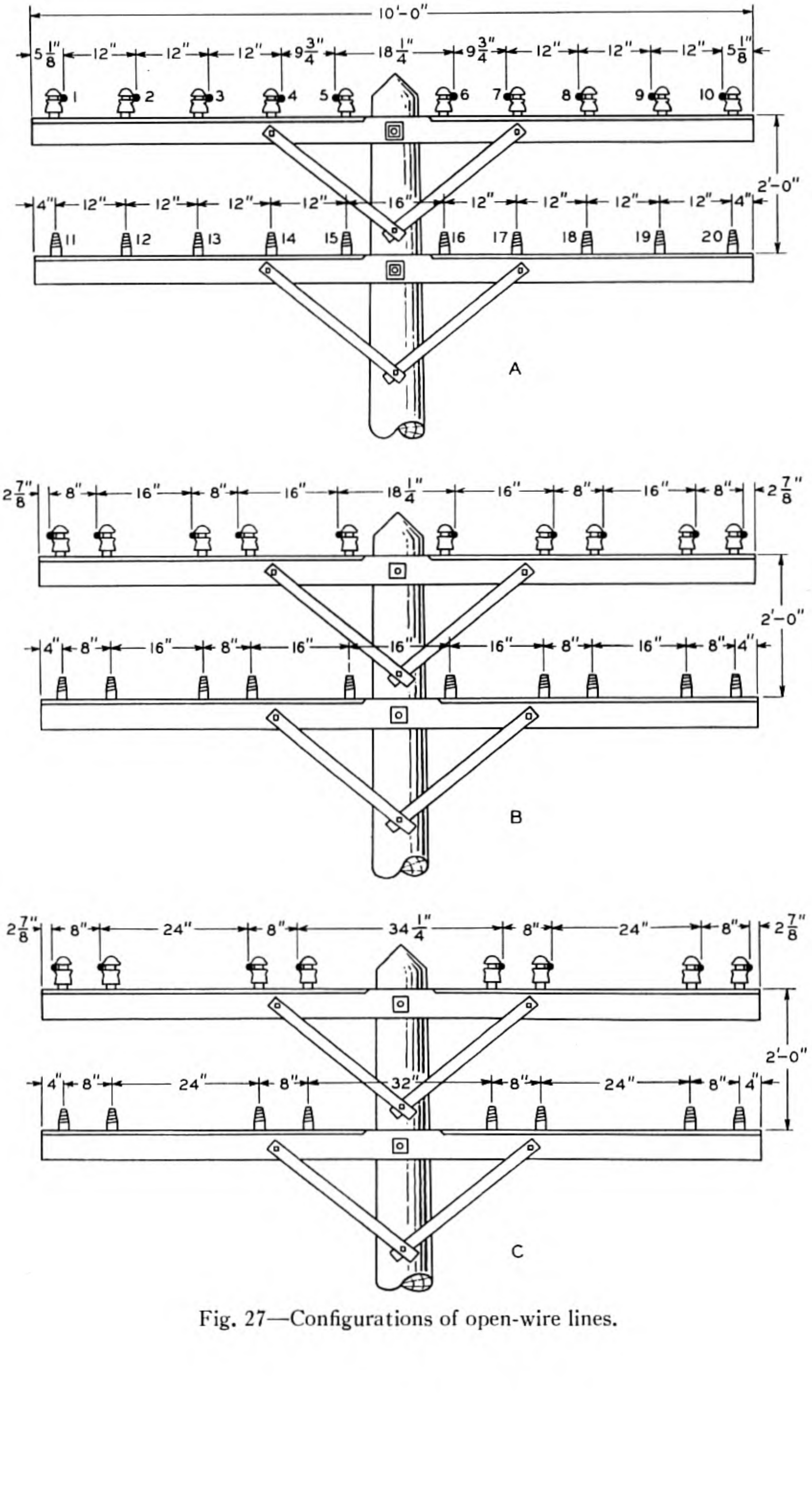

Fig. 27—Configurations of open-wire lines.

noise The "square phantom" indicated by *A* of Fig. 25 has theoretical possibilities but studies of the effect of wire spacing deviations make this arrangement appear impracticable.

The proposal to reduce the spacing of the wires of a pair from the historic value of 12 inches naturally raised the question of swinging contacts. However, extensive experience with 8-inch spacing has shown no appreciable increase in the number of wire contacts. This applies to lines where ordinarily the span length did not exceed about 150 feet. With long span crossings, crossarms were supported from steel strand at intervals of 260 feet or less.

The effectiveness of the reduction in wire spacing is indicated by the following table. The table shows the measured near-end and far-end crosstalk coefficients for important circuit combinations and for the two-pole head diagrams of Figs. 27-A and 27-B.

Crosstalk Per Mile Per Kilocycle—104-Mil Conductors

Pair Combination	Near-End Crosstalk		Far-End Crosstalk	
	12-Inch	8-Inch	12-Inch	8-Inch
1–2 to 3–4	974	439	74	34
3–4 to 7–8	133	47	77	15
1–2 to 11–12	653	326	66	30
1–2 to 13–14	40	18	58	24
3–4 to 13–14	549	288	155	69
1–2 to 21–22	163	78	35	16
1–2 to 23–24	55	28	43	17
3–4 to 23–24	107	55	75	36

General Transposition Design Methods

The preceding discussion will indicate that transposition design involves much more than consideration of the locations of the transpositions.

In practical design, the first step is to estimate the crosstalk due to unavoidable pole spacing and wire spacing irregularities for the configuration of wires under consideration and for a wide frequency range. This crosstalk represents the best that can be done with an ideal transposition design. It must be kept in mind that great precision is impracticable. The pole spacing of a line may change from time to time due to minor reroutings caused by highway changes, etc. The wire sag differences change with temperature and are affected by sleet.

If two long circuits are on adjacent or nearby pairs in one repeater section, they should, as far as practicable, be routed over non-adjacent pairs in other repeater sections in order to minimize the overall crosstalk between these two circuits. This crosstalk will usually be largely due to those parts of the parallel where the circuits are on adjacent or nearby pairs, since the pole line seldom has enough pairs to make it practicable to keep any two circuits far apart for a large proportion of the total parallel. It is important, therefore, to strive for the lowest possible crosstalk between adjacent or nearby pairs even though this requires permitting higher crosstalk between widely separated pairs than would otherwise be necessary.

For the adjacent or nearby pairs with naturally high crosstalk, limits on the type unbalance crosstalk are set which make this type of crosstalk small compared with that due to irregularities. Since the type unbalance crosstalk varies with frequency and, in general, increases with frequency, these limits are imposed only for the range of frequency which the line will be required to transmit. It is not advisable to go beyond this, since more severe limits require closer spacing of transpositions and the increased number of transpositions would make the "pole spacing" irregularity crosstalk larger. For the well-separated pairs with naturally lower crosstalk, the type unbalance crosstalk rather than the irregularity crosstalk may be allowed to control with the same idea in mind of requiring a minimum number of transposition points.

Fig. 28 indicates the method used generally in the Bell System for arranging transpositions with 32 transposition poles. The arrangements shown are called fundamental types. They are iterative, i.e., if the first two-interval length is transposed at the center, each following two-interval length is likewise transposed, etc. Various other arrangements called hybrid types are possible but in the long run there appears to be no advantage from their use except in the case of side circuits of phantoms. In this case the transposition pattern may change when the side circuit changes pin positions at a phantom transposition.

The fundamental types may be extended to involve 64, 128, 256, etc., transposition poles. Types involving 128 transposition poles are often used.

A long line, say 100 miles, is divided into short lengths called transposition sections. With the latest transposition designs, sections having 128, 64, 32, 16 and 8 transposition poles are provided. The nominal lengths of these sections vary from 6.4 to .25 mile. The purpose of these sections is to provide an approximate balance against crosstalk (and induction from power circuits) in short lengths and thus to allow for unavoidable discontinuities in the exposure between circuits such, for example, as points where circuits branch off the line. Transposition arrangements must be chosen for each circuit in each type of section to ensure this approximate crosstalk balance.

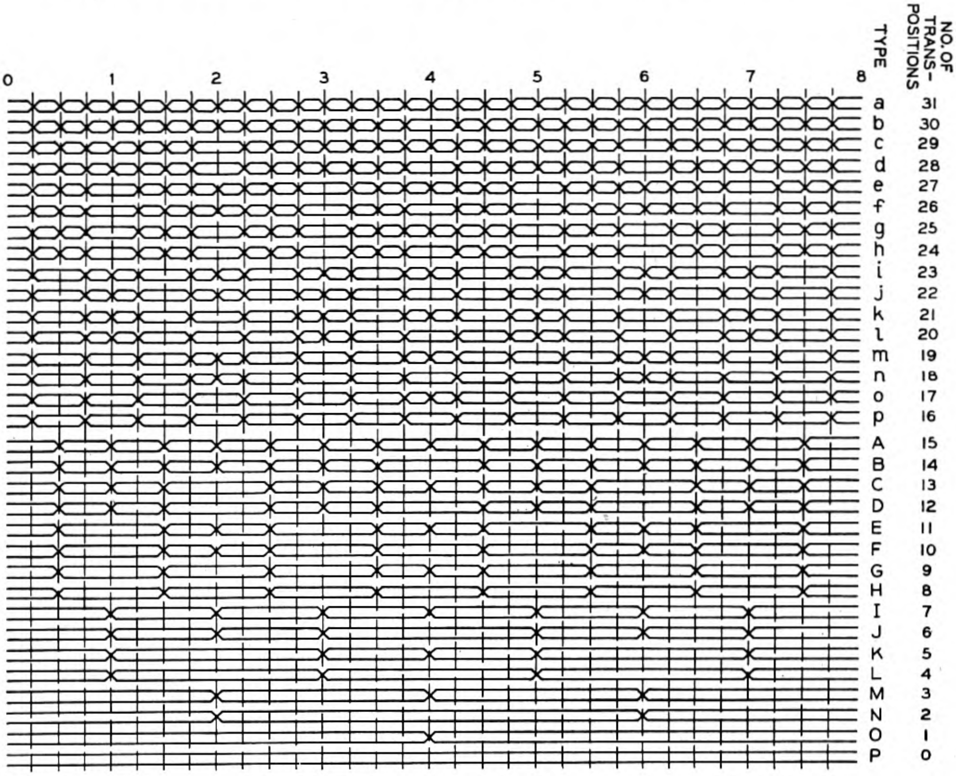

Fig. 28—Fundamental types for 32 transposition poles.

Certain lines have few, if any, discontinuities and a succession of the longest type of section is used. To improve the effectiveness of the transpositions, junction transpositions are used at the junctions of successive similar sections. For such lines it would be more effective to use longer transposition sections and not require that all circuits be approximately balanced in a short length. Such a special design would be impracticable, however, since it would be too inflexible in regard to circuit changes, etc.

In choosing the transposition arrangements for a section it must be kept in mind that the object is to meet certain crosstalk limits for a succession of sections considering both type unbalance and irregularity crosstalk. The method of procedure is discussed below.

Evolution of Transposition Designs

In designing transposition systems it must be kept in mind that much of the crosstalk is due to irregularities and is a matter of chance. Theoretically the crosstalk elements due to all of the various irregularities might chance to add directly. This is highly improbable and if the design were based on making this limiting condition satisfactory, the expense would be very great. Practically, therefore, the designs are based on exceeding a tolerable value a small percentage of the time. If, in practice, the tolerable value happens to be exceeded and this is not found to be due to an error in construction, the unfortunate adding up of crosstalk elements can be broken up by a different connection of circuits at the offices.

The tolerable values commonly chosen are 1000 crosstalk units (60 db) for open-wire carrier circuits and 1500 units (56 db) for voice-frequency open-wire circuits, which tend to have more line noise than cable or carrier circuits. These limits apply to the crosstalk between terminating test boards with the circuits worked at net losses of about 9 db.

Before proceeding with the design of the individual transposition sections which are but a few miles long, it is evidently necessary to determine what part of the overall limit can properly be assigned to an individual section. Assumptions must first be made as to typical and limiting lengths in which circuits are on the same pole line and in which adjacent or nearby circuits continue in this relation. A representative repeater layout must then be chosen. The repeater layout is very important, since the crosstalk in each repeater section is propagated to the circuit terminal and amplified or attenuated, depending on the arrangement of the repeaters. As a matter of fact, the layout of repeaters must be governed to a considerable extent by crosstalk considerations.

On the assumption that the relative magnitudes and phase relations of the crosstalk couplings in the various repeater sections are a matter of chance the tolerable crosstalk in a single repeater section can be estimated by the use of probability laws. Similarly the tolerable value for any part of the repeater section can be estimated. These probability methods apply very well to crosstalk due to irregularities. Type unbalance crosstalk is systematic, however, and in assigning tolerable values of type unbalance crosstalk in a transposition section, it is necessary to consider how the crosstalk values for various transposition sections may add up.

It is not likely that there will be systematic building up of type unbalance crosstalk in successive repeater sections and, therefore, the tolerable crosstalk per repeater section may be estimated by probability methods. The total of the irregularity crosstalk and the type unbalance crosstalk in a repeater section is a matter of chance and may be estimated from probability theory. Conversely, the part of the tolerable crosstalk which may be assigned to type unbalance crosstalk may be estimated. As noted above, the allowance for type unbalance crosstalk for adjacent or nearby pairs is usually made so small that irregularity crosstalk controls the total. The maximum permissible carrier frequency is, then, the frequency at which the irregularity crosstalk just reaches the tolerable value. Having determined tolerable values of type unbalance crosstalk for a repeater section for the various pair combinations, tolerable values for the individual transposition sections must be determined.

If a repeater section involves a number of different types of transposition sections it is not likely that there will be a systematic building up of type unbalance crosstalk. Factors are, therefore, worked out to relate the crosstalk in a succession of similar transposition sections to that in one section. Numerous factors are required since they depend upon the transpositions at the junctions of the sections. A study of such factors indicates values which it is reasonable to assign to an individual transposition section in order to avoid excessive type unbalance crosstalk in a complete repeater section.

In the case of a voice-frequency transposition system, both near-end and far-end type unbalance limits must be set. The far-end limits are usually easily met. In the case of a transposition system for carrier systems, far-end crosstalk is controlling and the far-end type unbalance limits are important. The "reflection crosstalk" previously discussed depends, however, on both the magnitude of the near-end crosstalk and on the impedance mismatches. Information on the degree to which it is practicable to reduce these mismatches must be available in order to set limits on near-end type unbalances at carrier frequencies.

Pairs used for carrier systems are usually also used for voice-frequency telephone systems and in designing transpositions for these pairs crosstalk limits suitable for both types of systems must be met. In practice, an existing line may have only a part of the pairs retransposed for carrier operation and in designing a system of transpositions for such retransposed pairs limits must be set for the crosstalk at voice frequencies between the retransposed pairs and the pairs not retransposed.

It has been the practice to transmit certain carrier telegraph frequencies in the opposite directions used for these frequencies in connection with carrier telephone, or, in some cases, program transmission circuits. At these frequencies near-end crosstalk limits must be set so as to limit the induced noise from the carrier telegraph.

When the type unbalance crosstalk limits are finally determined, the transposition designer must attempt to meet the requirements for all circuit combinations and all the transposition sections. It may be that the requirements can not be met and consideration must be given to modifications in the nature of the transmission systems. A vast amount of such preliminary transposition design work has been necessary in order to evolve the present transposition systems and transmission systems.

Such studies led to the development of non-phantomed circuits with 8-inch spacing since they indicated that multi-channel long-haul carrier operation on all pairs on a line was, in general, impracticable from the crosstalk standpoint with 12-inch phantomed pairs.

It may be noted that there are also difficulties in the crosstalk problem when 12-inch phantomed pairs are used for voice-frequency repeatered circuits. These circuits have a crosstalk advantage over carrier circuits in that the frequency is lower but they have an offsetting disadvantage in that they use the same frequency range in both directions. This makes the near-end crosstalk directly audible to the subscriber. As previously discussed the near-end crosstalk is inherently greater than the far-end crosstalk and, for this reason, practicable designs of multi-channel carrier systems do not allow near-end crosstalk to pass to the subscriber, the path being blocked by one-way amplifiers. While it takes fewer transpositions to control the type unbalance effects with voice-frequency transposition designs, for a given length of parallel the difficulties with crosstalk due to irregularities are about as great as with designs for multi-channel carrier operation.

The simple example of Fig. 29 illustrates the reasons for the difficulties with near-end crosstalk with the voice-frequency designs for 12-inch spaced pairs. It also illustrates the method of deducing the permissible crosstalk per repeater section as discussed above.

This figure indicates two paralleling repeatered circuits, each having six repeater sections of 10 db loss and five repeaters of 10 db gain. The net loss of each circuit is, therefore, 10 db. The near-end crosstalk values in the six sections are indicated by n_1 to n_6. The crosstalk coupling at A due to n_2 is just equal to n_2 since there is no net loss or gain in either circuit between A and B. There is also no net loss or gain between A and C, A and D, A and E or A and F. The total crosstalk coupling at A is, therefore, the vector sum of the six values

n_1 to n_6. If the crosstalk is due to irregularities the exact values of n can not be calculated but from the data collected on the crosstalk due to irregularities, the r.m.s. of all possible values may be estimated.

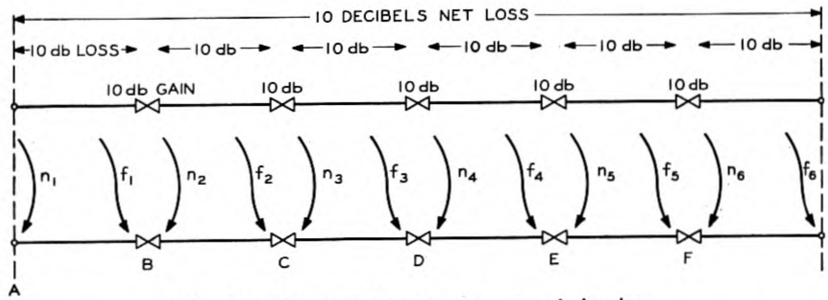

Fig. 29—Crosstalk between repeatered circuits.

Letting r_1 equal the r.m.s. value of n_1, etc., and using probability theory we may write:

$$R^2 = r_1^2 + r_2^2 + \cdots + r_6^2,$$

where R is the r.m.s. of all possible values of the near-end crosstalk at A. If $r_1 = r_2$, etc.

$$R = r\sqrt{6}.$$

The chance of the overall crosstalk deviating from R by any specified amount may be estimated by probability methods. It will be noted that the crosstalk in six repeater sections tends to be more severe than that in one section by $\sqrt{6}$ or, in other words, that the crosstalk varies as the square root of the length. If the use of repeaters were avoided by using more copper, for the same overall loss the crosstalk would be practically the same as with the repeatered circuits. *With the arrangement of repeaters shown* it is not the use of repeaters which causes the increase in crosstalk but rather the increase in circuit length without corresponding increase in circuit loss. For a given circuit length, circuit loss and wire size, other arrangements of repeaters may cause greater or less crosstalk.

If the repeaters of Fig. 29 are spaced farther apart, say 15 db instead of 10, there will be three line repeaters of 15 db gain each and terminal repeaters will be necessary to supply a terminal gain of 5 db in order to obtain a net loss of 10 db. The near-end crosstalk would be reduced by about $\sqrt{4} \div \sqrt{6}$ or 1.8 db because there are only four repeater sections but the terminal repeaters would amplify the near-end crosstalk by 5 db. The net increase would be 3.2 db. From the standpoint of near-end crosstalk, it is thus seen that close spacing between repeaters is very desirable.

In Fig. 29 the output-to-output far-end crosstalk in each repeater section is indicated by f_1 to f_6. The transmission path through any one of these crosstalk couplings is (for like circuits) a loss 10 db greater than the value of the coupling expressed as a db loss. With the repeater arrangement of the figure, the far-end crosstalk paths are attenuated by 10 db while the near-end crosstalk paths are not attenuated. Furthermore, the far-end crosstalk paths ordinarily introduce greater losses than the near-end paths. With greater spacing between repeaters, the near-end crosstalk is amplified but the far-end crosstalk (for like circuits) is still attenuated by the net loss of the circuits. At a given frequency the near-end crosstalk between such "two-wire" circuits is, therefore, much greater than the far-end crosstalk.

Review

Evidently the problem of keeping crosstalk between open-wire circuits within tolerable bounds is by no means a simple one. As we have seen, the work begins with consideration of complete circuits (telephone, program transmission or carrier telegraph) which may be hundreds or even thousands of miles long. The total crosstalk allowance for such long circuits must first be broken down into allowances for the various sections of line between repeaters and then into allowances for the individual transposition section, these individual sections ranging from less than 1/4 to about 6 miles in length.

Then bearing in mind that irregularities in pole spacing and in wire configuration set limits to crosstalk reduction which it is not practicable to overcome by transpositions, the crosstalk designer determines by computation whether, when considering these irregularity effects alone, the crosstalk requirements for the individual transposition sections can be met. If these requirements can not be met he must either have the general circuit layout altered so that, for example, the repeater gains will be more favorably disposed from the standpoint of crosstalk, or he must alter the pole head configuration so that the electrical separation between the circuits will be increased.

Having obtained an overall circuit layout and a configuration of the wires which makes it possible to attain the desired overall crosstalk results, the design of the transpositions proper is undertaken. In this work the transposition designer makes every effort to keep the number of transpositions at a minimum. He does this partly to save money but more particularly because he recognizes that more than enough transpositions do harm rather than good by increasing the number of pole spacing irregularities.

In dealing with the problems of crosstalk coupling between open-wire circuits, consideration must be given not only to the direct effect of one circuit on another but also to the indirect effect of the other circuits on the line. What happens is that the disturbing circuit crosstalks not only directly into the circuit under consideration but also into the group of other circuits and thence into the disturbed circuit. The name "tertiary" circuit has been given to this group of circuits although it is not in reality one circuit but rather any or all of the possible circuits which may be formed of the different wires. The system of transpositions must, therefore, not only substantially balance out the direct couplings between disturbing and disturbed circuits but must also substantially balance the couplings from the disturbing circuit into the "tertiary" circuit and from this "tertiary" circuit into the disturbed circuit.

Reflections of the electrical waves also add interest and complexity to the problem. Such reflections tend to increase crosstalk because the electrical waves which are changed in direction as a result of reflections crosstalk differently, and in many cases more severely, into neighboring circuits than do the waves traveling in the normal direction. The most important reflections occur at junctions between lines and office apparatus. The possibility of other reflections must also be considered, however, at intermediate points in the line which might be caused by inserted lengths of cable, change in spacing of wires, etc.

In working out the transposition designs, the fact that crosstalk between two paralleling circuits tends to manifest itself at both ends is of great importance. At the "near end" crosstalk coming from the disturbed circuit in a direction opposite to the transmission in the disturbing circuit must be considered. At the "far end" crosstalk coming in the same direction as the transmission in the disturbing circuit must be considered.

For telephone circuits which use the same path for transmission in both directions, the "near-end" crosstalk is considerably more severe than the "far-end" for two reasons: (1) The crosstalk per unit length of the paralleling circuits is greater; (2) the gains of the repeaters especially augment the "near-end" crosstalk. Voice-frequency open-wire telephone circuits have always been worked on this "one-path" basis and are good examples of circuits in which "near-end" crosstalk is controlling and must be given principal consideration in working out transposition designs.

In the case of carrier circuits, it was found early in the development that if these circuits were worked on a one-path basis, the crosstalk would be prohibitively great. Consequently, carrier circuits are now designed to operate on a two-path basis. Two separate bands of frequencies are set aside, each being restricted, by means of one-way amplifiers and electrical filters, to transmission in one direction only. Each telephone circuit is then made up of two oppositely directed channels, one in each frequency band. Thus, direct "near-end" crosstalk is kept from passing to the telephone subscribers. Consequently, the "near-end" type of crosstalk needs to be considered only with respect to that portion which arises from electrical waves reflected at discontinuities in the circuits, which effects have already been mentioned.

In practice a pole line may have some of the pairs very frequently transposed to make them suitable for carrier frequency operation and other pairs less frequently transposed and suitable only for voice-frequency operation. A system of transpositions must permit any arrangement of the two types of pairs which may be found economical for a given line and layout of circuits. Each pair must meet limits for near-end and far-end crosstalk to any other pair which may crosstalk into it in its frequency range. Pairs used only for voice frequencies are usually phantomed and transpositions must, of course, be designed for the phantom circuits as well as the side circuits. The design of a transposition system is, therefore, extraordinarily complicated and tedious and, to paraphrase the Gilbert and Sullivan policeman, "A transposer's lot is not a happy one."

Bibliography

The published material on the matter of open-wire crosstalk and transposition design appears to be very limited. The following papers are of interest:

The Design of Transpositions for Parallel Power and Telephone Circuits, H. S. Osborne. *Trans. of A. I. E. E.*, Vol. XXXVII, Part II, 1918.

Telephone Circuits with Zero Mutual Induction, Wm. W. Crawford, *Trans. of A. I. E. E.*, Vol. XXXVIII, Part I, 1919.

Measurement of Direct Capacities, G. A. Campbell. *Bell System Technical Journal*, July, 1922.

Propagation of Periodic Currents Over a System of Parallel Wires, John R. Carson and Ray S. Hoyt. *Bell System Technical Journal*, July, 1927.

On Crosstalk Between Telephone Lines, M. Vos. L. M. Ericsson Review, English Edition, 1930, Vol. 7.

Application of High Frequencies to Telephone Lines, M. K. Kupfmuller. Presented Before International Electrical Congress, July, 1932.

Probability Theory and Telephone Transmission Engineering, Ray S. Hoyt. *Bell System Technical Journal*, Jan., 1933.

APPENDIX

Calculation of Crosstalk Coefficients

This appendix will first cover methods of calculating the coefficients of transverse crosstalk coupling. It is necessary to calculate both near-end and far-end crosstalk coefficients which involve both direct and indirect components of transverse crosstalk coupling. Coefficients for the direct and for the indirect components will be derived separately and then combined to obtain the total coefficients.

Ordinarily, the indirect effect cannot be readily computed with good accuracy and the total coefficients are usually measured. As previously noted, the method of computing the indirect effect can be used with fair accuracy, however, and it is useful in cases where measurements are impracticable.

The crosstalk between frequently transposed circuits may be calculated with the aid of the above coefficients of transverse coupling and in addition an "interaction crosstalk coefficient" relating to interaction crosstalk coupling of the most important type. The relation of this interaction coefficient to the far-end coefficient of transverse coupling is also discussed herein.

Direct Crosstalk Coefficients

Figure 30 indicates the definitions of the direct crosstalk coefficients used in computing the direct component of the transverse crosstalk coupling. This figure shows a thin transverse slice in a parallel between two long circuits a and b, the thickness of the slice being the

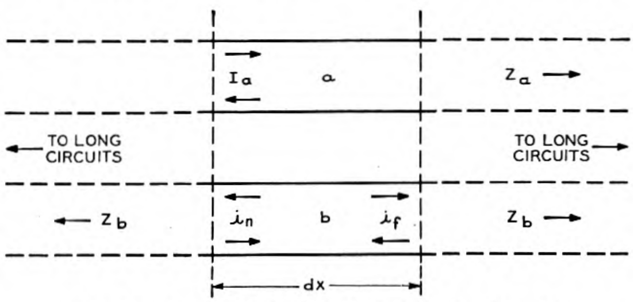

Fig. 30—Crosstalk in a single infinitesimal length.

infinitesimal length dx. Circuit a is energized from the left, the current entering dx being I_a. Propagation of I_a through dx results in near-end and far-end currents i_n and i_f in circuit b at the ends of dx. Since the coefficients are the crosstalk per mile per kilocycle, the near-end coefficient N and the far-end coefficient F may be expressed as follows:

$$N = \text{limit of } \frac{i_n}{I_a} \cdot \frac{10^6}{Kdx} \text{ as } dx \text{ approaches zero.}$$

$$F = \text{limit of } \frac{i_f}{I_a} \cdot \frac{10^6}{Kdx} \text{ as } dx \text{ approaches zero.}$$

where K is the frequency in kilocycles. For circuits of different characteristic impedances Z_a and Z_b the above current ratios should be multiplied by the square root of the ratio of the real parts of Z_b and Z_a. This correction is not included in the expressions for N and F derived below.

Figure 31 indicates the equivalent electromotive forces which, if impressed on the disturbed circuit b, would cause the same direct crosstalk currents as the electric and magnetic fields of the disturbing circuit. The series and shunt electromotive forces V_m and V_e correspond to the magnetic and electric components of the field and cause crosstalk current i_m and i_e. These currents are about equal in magnitude and they add almost directly at the near end of the length dx and subtract almost directly at the far end. The near-end coefficient is, therefore, inherently much greater than the far-end coefficient.

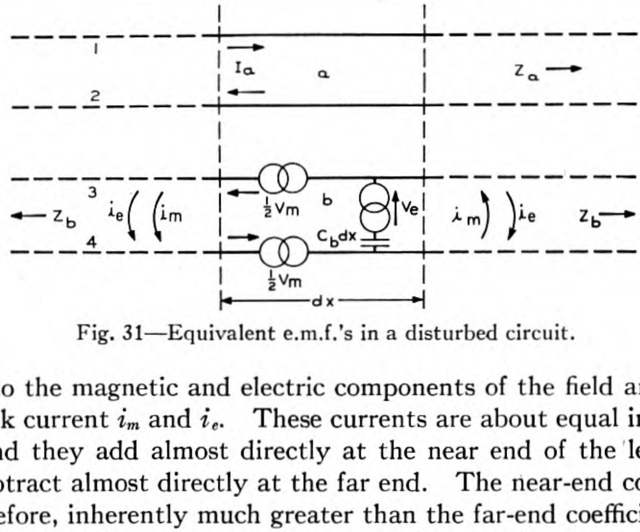

Fig. 31—Equivalent e.m.f.'s in a disturbed circuit.

To calculate i_e, the crosstalk current due to the electric field of circuit a, it is necessary to know the shunt voltage V_e. This depends on the charges on the wires of circuit a in the length dx. These charges are due to a voltage V impressed on the left-hand end of circuit a which may be remote from the length dx. Since it is desired to transmit on the metallic circuit a and not on the circuit composed of its wires with ground return, care is taken to "balance" the impressed voltage, i.e., this sending circuit has equal and opposite voltages between its two sides and ground with circuit a disconnected.

The impressed voltage V is propagated to the left-hand end of dx. Letting V_a be the voltage across circuit a at this point, it will be shown that V_a would be balanced except for the effect of interaction crosstalk which is excluded from consideration for the present. Designating the wires of circuit a as 1 and 2, the balanced voltage V_a causes charges Q_1 and Q_2 per unit length on these wires in the length dx. These charges are affected by the presence of other wires in the length dx and they are usually unbalanced. There will be equal and opposite or balanced charges $\pm \frac{Q_1 - Q_2}{2}$ on each wire and unbalanced equal charges $\frac{Q_1 + Q_2}{2}$ on each wire. Since the direct crosstalk is defined as the effect of balanced charges and currents, only the balanced charges should be considered in computing V_e. Letting $Q_a = \frac{Q_1 - Q_2}{2}$ or the balanced charge on wire one per unit length, then:

$$Q_a = V_a C_a = I_a Z_a C_a,$$

where C_a is equal to the "transmission capacitance" per mile, i.e., the capacitance used in calculating α_a the attenuation constant and Z_a the characteristic impedance of circuit a.

The above expression for Q_a includes the reaction of charges in the disturbed circuit. This reaction should not theoretically, be included at this time, since, for convenience in calculation, the disturbed circuit is assumed to have the impressed voltages V_m and V_e but no crosstalk currents or charges as yet. The effect on Q_a of charges in the disturbed circuit, is, however, usually small compared with the effect of charges in various tertiary circuits.

Designating the conductors of circuit b as 3 and 4, V_e is the difference of the potentials of the electric field at 3 and 4 caused by the balanced charges per unit length on 1 and 2. Therefore:

$$V_e = V_3 - V_4 = (Q_a p_{13} - Q_a p_{23}) - (Q_a p_{14} - Q_a p_{24}),$$

where p_{13}, etc., are the potential coefficients.

For c.g.s. elst. units, $p_{13} = 2 \log \frac{s_{13}}{r_{13}}$ where s_{13} and r_{13} are the distances indicated by Fig. 32. Therefore:

$$V_e = V_a C_a (p_{13} - p_{23} - p_{14} + p_{24}) = V_a C_a p_{ab}.$$

The capacitance C_a may be obtained from measurements on a short length of a multi-wire line. Its value is, however, only a few per cent greater than C_a' the value for a single pair line (without capacitance at the insulators). For a single pair having like wires in a horizontal

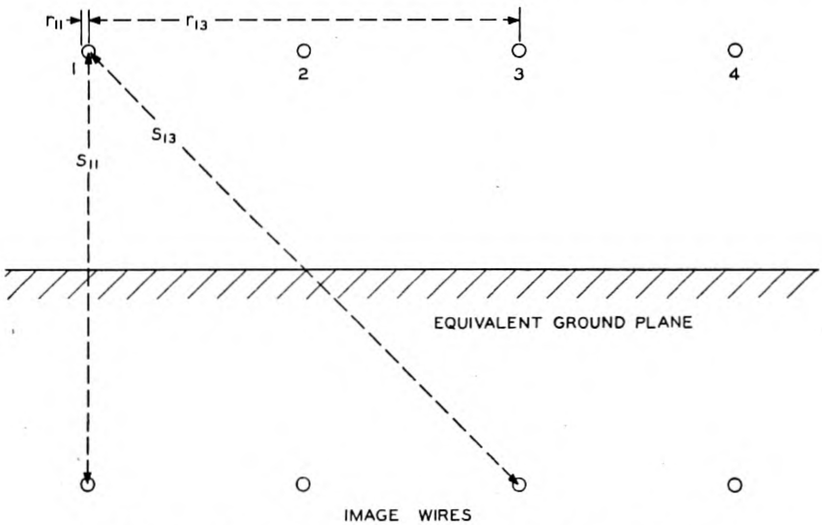

Fig. 32—Distances used in computing potential coefficients.

plane, C_a' is readily calculated as follows:

$$C_a' = \frac{1}{2(p_{11} - p_{12})},$$

where p_{11} in c.g.s. elst. units is:

$$2 \log_e \frac{s_{11}}{r_{11}}.$$

The distances s_{11} and r_{11} are indicated on Fig. 32.

The expression for V_e may be written:

$$V_e = V_a C_a p_{ab} = V_a C_a' p_{ab} \frac{C_a}{C_a'} = V_a T_{ab} \frac{C_a}{C_a'}.$$

The coefficient T_{ab} is called the "voltage transfer coefficient." It is readily computed since it is a function of potential coefficients and it is independent of the system of units used in computation. Since C_a is about equal to C_a', T_{ab} is about equal to the ratio of V_e to V_a.

The shunt voltage V_e drives a current through the shunt admittance of circuit b in the infinitesimal length dx of Fig. 31. This shunt admittance is $(G_b + j\omega C_b)dx$ which is very nearly equal to $j\omega C_b dx$ where C_b is the transmission capacitance per mile of circuit b and

$\omega = 2\pi f$ where f is the frequency in cycles per second. This current divides equally between the two ends of circuit b. The near-end current is:

$$i_e = -\frac{1}{2}\frac{V_e}{\dfrac{1}{j\omega C_b dx} + \dfrac{Z_b}{2}}.$$

The near-end direct crosstalk coefficient due to the electric field of circuit a may be called N_e and is the limiting value of the following expression as dx approaches zero:

$$\frac{i_e}{I_a}\cdot\frac{10^6}{Kdx} = -\frac{1}{2KI_a}\cdot\frac{V_e 10^6}{\dfrac{1}{j\omega C_b} + \dfrac{Z_b}{2}dx},$$

where K is the frequency in kilocycles. The near-end direct crosstalk coefficient for the electric field is, therefore:

$$(1)\quad N_e = -\frac{V_e 10^6 j\omega C_b}{2KI_a} = -\frac{Z_a T_{ab} j\omega C_b 10^6}{2K}\cdot\frac{C_a}{C_a'}$$

$$= -j\pi Z_a T_{ab} C_b 10^9 \frac{C_a}{C_a'}.$$

The far-end current due to V_e of Fig. 31 is $-i_e$ and, therefore, the far-end coefficient due to the electric field is $-N_e$.

The near-end and far-end crosstalk currents of Fig. 31 due to the magnetic field are alike and are designated i_m which may be calculated as follows:

$$i_m = \frac{V_m}{2Z_b} = -\frac{I_a j\omega M_{ab} dx}{2Z_b}.$$

The near-end or far-end crosstalk coefficients for the magnetic field may be called N_m and F_m. They are alike and equal to the limit of:

$$\frac{i_m}{I_a}\cdot\frac{10^6}{Kdx} \text{ as } dx \text{ approaches zero.}$$

Therefore:

$$N_m = F_m = -\frac{j\omega M_{ab}}{2Z_b K}10^6 = -\frac{j\pi M_{ab}}{Z_b}10^9.$$

In the above, M_{ab} is the mutual inductance per unit length between circuits a and b. It is calculated in the same manner as p_{ab} used in computing V_e. These methods of computing V_e and V_m from the distances r_{13}, r_{14}, s_{13}, etc., of Fig. 32 are not precise but are sufficiently accurate for open-wire circuits since the diameters of the wires are small compared with their interaxial distances. The "image" wires of Fig. 32 should, theoretically, be located farther below the equivalent ground plane for calculations of mutual inductance. This alters s_{14}, etc. Since the distances between wires are small compared to those between wires and images, the values of s are all about equal and have practically no effect on the value of p_{ab}.

Therefore, M_{ab} in c.g.s. elmg. units may be assumed numerically equal to p_{ab} or:

$$M_{ab} = p_{ab} = \frac{T_{ab}}{C_a'}.$$

In c.g.s. elst. units $C_a' = \dfrac{1}{2(p_{11} - p_{12})}$ which is also the expression for $1/L_a'$ in c.g.s. elmg. units where L_a' is the *external* inductance of circuit a, i.e., the inductance due to the magnetic field external to the wires of circuit a. Therefore:

$$M_{ab} = T_{ab}L_a',$$

where M_{ab} and L_a' may be expressed in any system of units.

The near-end or far-end direct coefficient for the magnetic field may, therefore, be written:

$$N_m = F_m = -\frac{j\pi T_{ab}L_a'}{Z_b}10^9. \tag{2}$$

The above expression is almost equal to N_e, the near-end coefficient for the electric field.

It may be written:

$$N_m = N_e\left[\frac{j\omega L_a' j\omega C_a'}{Z_a j\omega C_a Z_b j\omega C_b}\right].$$

Now $Z_a j\omega C_a$ is very nearly equal to $Z_a(G_a + j\omega C_a)$ which is γ_a. Likewise $Z_b j\omega C_b$ is very nearly equal to γ_b. If the circuits had no resistance or leakance the propagation constant would be $\gamma_0 = j\omega\sqrt{L_a'C_a'}$ or $j\omega/v$ where v is the speed of light in miles per second. Therefore:

$$N_m = \frac{N_e\gamma_0^2}{\gamma_a\gamma_b} \text{ very nearly.}$$

The total direct crosstalk coefficients are:

$$N_d = N_m + N_e = N_e\left(1 + \frac{\gamma_0^2}{\gamma_a\gamma_b}\right) = 2N_e \text{ approx.} \tag{3}$$

At carrier frequencies the ratio of γ_0 to γ_a (or γ_b) is about equal to the ratio of the actual speed of propagation to the speed of light, i.e., 180,000 to 186,000 or about .97. Therefore $\left(1 + \frac{\gamma_0^2}{\gamma_a\gamma_b}\right)$ is about 1.94.

$$(4)\quad F_d = N_m - N_e = N_e\left(\frac{\gamma_0^2}{\gamma_a\gamma_b} - 1\right)$$

$$= N_e\left(-.06 + j\frac{90}{\pi}\frac{\alpha_a + \alpha_b}{K}\right) \text{ approx.}$$

The attenuation of the disturbing and disturbed circuits may not be neglected in evaluating the expression $\left(\frac{\gamma_0^2}{\gamma_a\gamma_b} - 1\right)$.

The expression given for N_e in equation (1) above may be written:

$$N_e = -\frac{\gamma_a T_{ab} 10^6}{2K}\frac{C_a}{C_a'}\frac{C_b}{C_a}.$$

This assumes $Z_a j\omega C_a$ equals γ_a. At carrier frequencies γ_a is about equal to $j\beta_a$ which is about $j\pi K/90$ since the speed of propagation is about 180,000 miles per second. The expression for N_e may, therefore, be written in the following simple approximate form:

$$N_e = -j\frac{\pi T_{ab}10^6}{180}\frac{C_a}{C_a'}\cdot\frac{C_b}{C_a}.$$

The ratio of C_a to C_a' does not ordinarily exceed 1.02. For like circuits, therefore:

$$N_e = \frac{-j\pi T_{ab}10^6}{180} \text{ approx.}$$

On Fig. 33 the magnitudes of N_d and F_d are plotted against frequency for 8-inch spaced conductors .128-inch in diameter. Both F_d and N_d are divided by T_{ab} to make the curves applicable to any circuit combination. These curves show that N_d is practically independent of frequency (above a few hundred cycles) but F_d decreases rapidly with frequency for several thousand cycles.

Indirect Crosstalk Coefficients

Expressions for the indirect crosstalk coefficients used in computing the indirect component of transverse crosstalk coupling will now be derived. The derivation first covers the case of a single representative tertiary circuit. Fig. 34 shows a thin transverse slice of the parallel of the three circuits, the thickness of the slice being the infinitesimal length dx. The only tertiary circuit to be considered for the present is the metallic circuit composed of wires 5 and 6 and designated as c. There are other possible tertiary circuits in the system of 6 wires, for example, the phantom circuit composed of wires 1 and 2 as one side and 5 and 6 as the other side. The method of estimating the total effect of all possible tertiary circuits will be discussed later.

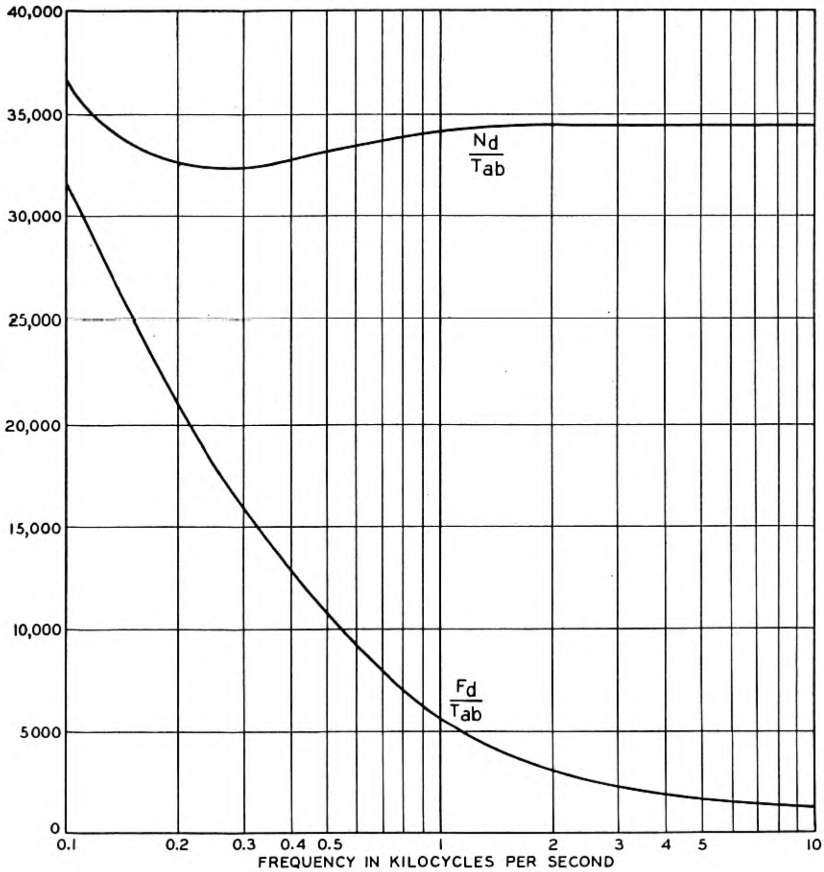

Fig. 33—Variation of direct crosstalk coefficients with frequency.

The immediate problem is to compute the crosstalk currents in circuit b at the ends of the length dx due to currents and charges in circuit c in this length and caused by transmission over circuit a through dx.

The crosstalk currents in circuit b due to currents and charges in circuit a were computed by determining the equivalent series and shunt e.m.f.'s in circuit b. The effect of currents and charges in circuit c on crosstalk currents in circuit b may be computed in a similar manner. The series e.m.f. in circuit b proportional to the current in circuit c will, however, be negligible compared with the series e.m.f. proportional to current in circuit a. This is evident since the current in circuit c is a crosstalk current which approaches zero as dx approaches zero while the current in circuit a does not vary with dx.

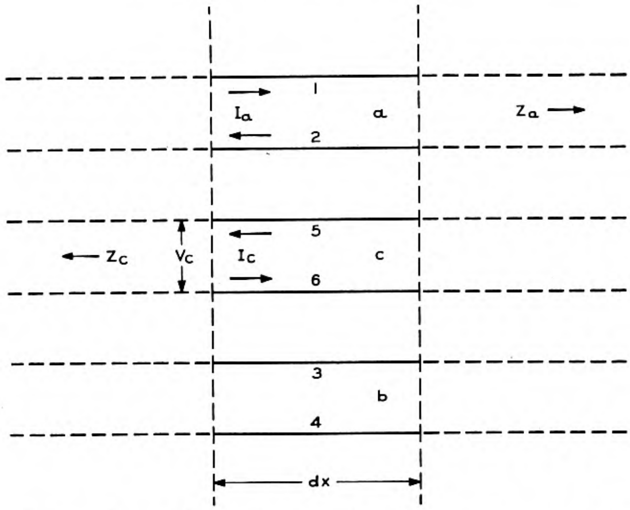

Fig. 34—Schematic used in deriving formulas for indirect crosstalk coefficients.

The shunt e.m.f. in circuit b dependent on the charges of circuit c is not, however, negligible compared with the shunt e.m.f. in circuit b due to charges in circuit a since the charges in both a and c approach zero as dx decreases. In other words the magnetic field of circuit c may be neglected but the electric field must be considered. (Both fields must be considered in computing interaction crosstalk.)

To determine the equivalent shunt e.m.f. in circuit b which depends upon the electric field of circuit c the voltage between the wires of circuit c must be determined. If circuit c did not exist, the electric field of circuit a would cause a difference of potential between the points actually occupied by wires 5 and 6 at the left-hand end of dx in Fig. 34. This difference of potential would be:

$$V_{ac} = V_a C_a' p_{ac} = V_a T_{ac}.$$

With circuit c present, this difference of potential is changed to V_c, the actual voltage across circuit c. The voltage could not change from V_{ac} to V_c without charges on circuit c and the charge per wire per unit length is proportional to the change in voltage from V_{ac} to V_c which may be designated U_c. The equivalent shunt e.m.f. in circuit b due to the presence of charges in circuit c is, therefore, proportional to U_c.

By definition:

$$V_{ac} + U_c = V_c \quad \text{or} \quad U_c = V_c - V_{ac}.$$

Since the crosstalk current in circuit c approaches zero as dx approaches zero, V_c must also approach zero and U_c approaches $- V_{ac}$. The shunt e.m.f. in circuit b due to charges on circuit a was computed as:

$$V_e = V_a T_{ab} \frac{C_a}{C_a'} \cdot$$

To allow for the *electric* field of circuit c, V_e must be augmented by:

$$V_e' = U_c T_{cb} \frac{C_c}{C_c'} = - V_{ac} T_{cb} \frac{C_c}{C_c'} = - V_a T_{ac} T_{cb} \frac{C_c}{C_c'} \cdot$$

Since the part of the direct near-end crosstalk coefficient resulting from V_e was found to be $N_e = - j\pi Z_a T_{ab} C_b 10^9 \frac{C_a}{C_a'}$, by proportion the indirect near-end coefficient resulting from V_e' will be:

$$N_i = j\pi Z_a T_{ac} T_{cb} C_b 10^9 \frac{C_c}{C_c'} = \frac{j\pi T_{ac} T_{cb} 10^6}{180} \text{ approx.} \tag{5}$$

Since the far-end crosstalk current resulting from a shunt voltage in circuit b is opposite in sign to the near-end current, the indirect far-end coefficient will be:

$$F_i = - N_i. \tag{6}$$

Total Crosstalk Coefficients

The total near-end and far-end crosstalk coefficients used in computing transverse crosstalk coupling will be the sum of the direct and indirect coefficients or:

$$N = N_d + N_i. \tag{7}$$

$$F = F_d + F_i = F_d - N_i. \tag{8}$$

The expressions for F_i and N_i are about independent of frequency in the carrier-frequency range because Z_a does not depend much on frequency above a few thousand cycles, C_b is about independent of frequency and $T_{ac} T_{cb}$ depends only on the cross-sectional dimensions of the wire configuration.

Since, as indicated by Fig. 33, N_d is usually about independent of frequency and since $N = N_d + N_i$ is largely determined by N_d, the near-end coefficient N is about independent of frequency above a few hundred cycles. The far-end coefficient F is about independent of frequency above a few thousand cycles where it is largely determined by F_i.

The preceding discussion of indirect crosstalk coefficients covered only the effect of charges in the single metallic tertiary circuit c of Fig. 34. The indirect coefficient in a practical case may be estimated with fair accuracy by considering all the more important tertiary circuits in a similar manner. It was shown that the final voltage of tertiary circuit c was zero. Similarly, the final voltage of each tertiary circuit is zero. This includes any tertiary circuits involving the two wires of the disturbing circuit in multiple. The average voltage of the two wires of the disturbing circuit is zero and the voltage across the disturbing circuit is balanced. As previously stated, this voltage does not become unbalanced as a result of transverse crosstalk in any infinitesimal length but it may become unbalanced due to interaction crosstalk.

The charges per unit length on the various tertiary circuits are the same as those which would be caused by impressing a system of voltages equal and opposite to those induced by the balanced charges per unit length which would be on the two wires of the disturbing circuit if this circuit were the only pair on the line. Assuming such a system of impressed voltages, it is not practicable to accurately compute the charges in any tertiary circuit since this depends on the voltages impressed on all the tertiary circuits and the couplings between the various tertiary circuits. Advantage may be taken, however, of the fact that the charge on a tertiary circuit will depend mostly on the voltage impressed on that circuit provided it is not heavily coupled with other circuits.

It is possible to divide the various voltages impressed on the tertiary circuits into components such that (1) equal voltages are impressed on wires of a "ghost" circuit composed of all the wires on the line with ground return, (2) balanced voltages are impressed on each pair used for transmission purposes (except the disturbed and disturbing circuits) and (3) balanced voltages are impressed on each possible phantom of two pairs used for transmission purposes.

Such a system of impressed voltages and tertiary circuits is convenient for computation since the charge on any tertiary circuit largely depends on the voltage impressed on that circuit. If accurate calculations of the charges were practicable, a simpler system of tertiary circuits could be used to obtain the same final result, i.e., single-wire tertiary circuits with ground return could be used. Computation with such tertiary circuits is impracticable because of the large coupling between them.

In the elaborate system of tertiary circuits described above, the ghost circuit may be neglected. The voltage impressed across this circuit is the average of all the voltages impressed on the various wires. These voltages may be plus or minus and the average tends to be small. Also, the charge per pair per impressed volt is usually much less for the ghost circuit than for a phantom circuit due to the relatively small capacitance between a pair and ground as compared with that between two pairs.

The pairs used for transmission purposes may usually be disregarded, also, since their coupling with the disturbing and disturbed circuits is much smaller than that of the phantom tertiary circuits.

The practical method of computing the indirect crosstalk coefficient is, therefore, to consider as tertiary circuits a considerable number of phantoms composed of pairs used for transmission purposes including the disturbed and disturbing pairs. In calculating the charge in any tertiary circuit, the voltages impressed on other tertiary circuits are disregarded.

In calculating the effect of a single tertiary circuit c, the expression for the indirect coefficient contained the factor $T_{ac}T_{cb}$. To estimate the effect of all the tertiary circuits, this factor should be replaced by:

$$\frac{2}{m}\sum T_{ap}T_{pb}.$$

This expression assumes that there are n pairs on the line and that $m - 2$ of these pairs are close enough to the disturbing and disturbed pairs to appreciably affect the indirect crosstalk between them. The subscript p indicates any phantom of the m pairs including the disturbed and disturbing pairs. The summation is for all possible phantoms each consisting of two of the m pairs. If the voltages induced by the balanced charges Q_a' of pair a are V_r and V_s for the two sides of a phantom, the balanced voltage assumed to be impressed across the phantom is $\frac{2}{m}(V_s - V_r)$. Other parts of V_r and V_s are used in the "ghost" voltage and in balanced voltages across other phantoms.

T_{ap} and T_{pb} are voltage transfer coefficients relating balanced impressed voltage on the disturbing circuit to induced voltage on the disturbed circuit. T_{ap} involves C_a' the transmission capacitance of circuit a on a single pair line. T_{pb} involves the transmission capacitance of a particular phantom on a line having only that phantom present. This capacitance is the ratio of balanced charge (on each side of the phantom) to the balanced impressed voltage. The phantom capacitance may be readily estimated from the potential coefficients. For example, if the phantom involves pairs 1–2 and 5–6 the phantom capacitance is very nearly:

$$C_p = \frac{4}{p_{11} + p_{22} + p_{55} + p_{66} + 2p_{12} + 2p_{56} - 2p_{15} - 2p_{25} - 2p_{16} - 2p_{26}}.$$

If the disturbing circuit is pair 1–2 and the disturbed circuit is pair 3–4:

$$T_{ap} = \frac{p_{15} + p_{16} - p_{25} - p_{26}}{2(p_{11} - p_{12})}$$

$$T_{pb} = \frac{C_p}{2}(p_{13} + p_{23} + p_{45} + p_{46} - p_{14} - p_{24} - p_{35} - p_{36}).$$

These computations of indirect coefficients are necessarily laborious. They can be simplified to some extent by ignoring phantoms for which either T_{ap} or T_{pb} is zero or small. For example, the voltage transfer coefficient is zero for pair 1–2 to such phantoms as 1–2 and 11–12, 11–12 and 21–22, etc.

In the following table are given comparisons of far-end crosstalk coefficients as measured in a 40-wire line and as computed by the methods discussed above. The spacing of the various wires and crossarms is indicated by Fig. 27A. The measured values are for 40 wires and the computed values are for 10, 20 and 30 wires. It will be seen that a considerable number of wires must be taken into account in the computations in order to obtain a fair check with the coefficient measured for a heavy line.

Far-End Crosstalk Per Mile Per Kilocycle

	Combination		
	1–2 to 3–4	1–2 to 11–12	1–2 to 9–10
Computed for 10 wires	45	28	
Computed for 20 wires	63	47	11
Computed for 30 wires	69	58	22
Measured for 40 wires	74	70	21

Interaction Crosstalk Coefficient

It was assumed in the discussion of crosstalk coefficients that the "interaction crosstalk coefficient" $N_{ac}N_{cb}10^{-6}$ was nearly equal to $-2F_i\gamma_c/K$. This relation is deduced below, for a representative tertiary circuit c, from the expressions for F_i and N_d given by equations (3) and (6) above. N_{ac} may be obtained by using the expressions for N_e and N_d given by equations (1) and (3) above. In these equations, subscript c should be substituted for subscript b. The expression for N_{ac} becomes:

$$N_{ac} = -j\pi Z_a T_{ac} C_c 10^9 \frac{C_a}{C_a'}\left[1 + \frac{\gamma_0^2}{\gamma_a\gamma_c}\right].$$

Deriving a similar approximate expression for N_{cb}:

$$N_{ac}N_{cb}10^{-6} = -\frac{F_i}{2K}\gamma_c\frac{C_a}{C_a'}\left[1 + \frac{\gamma_0^2}{\gamma_a\gamma_c}\right]\left[1 + \frac{\gamma_0^2}{\gamma_c\gamma_b}\right].$$

This assumes $Z_c j\omega C_c = Z_c(G_c + j\pi C_c)$ which is γ_c.

Crosstalk measurements indicate that the ratio of γ_0 to γ_a or γ_b or γ_c is about .97 at carrier frequencies and C_a/C_a' about 1.02. Therefore:

$$N_{ac}N_{cb}10^{-6} = -\frac{F_i\gamma_c}{2K}1.02(1.94)^2 = -2F_i\gamma_c/K \text{ approx.}$$

The above discussion covers the case of a single tertiary circuit c. In the practical case the known crosstalk coefficient F_i includes the effect of a large number of tertiary circuits which have various values of γ_c. Obtaining the interaction crosstalk coefficient from the expression $-2F_i\gamma_c/K$ involves assuming a representative value of γ_c. At carrier frequencies γ_c is about equal to $j\beta_c$ which for all important tertiary circuits is in the neighborhood of $j\pi K/90$.

A known value of the crosstalk coefficient F_i expresses the effects of the electric fields of many tertiary circuits in one infinitesimal slice and includes the alteration of the field of any one tertiary circuit due to the presence of the others. The interaction crosstalk coefficient involves consideration of both electric and magnetic fields. In any one slice, the electric field of a tertiary circuit determines its crosstalk into the disturbed circuit but the current in the tertiary circuit at the end of the slice is determined by both the electric and magnetic fields of the disturbing circuit. The tertiary circuit current is transmitted into another slice and sets up electric and magnetic fields which both contribute to the crosstalk current in the disturbed circuit.

In the case of a single tertiary circuit, the electric and magnetic effects expressed by the interaction coefficient are simply related and the interaction coefficient is simply related to F_i. With many tertiary circuits the relation between the electric and magnetic fields of any one tertiary circuit is altered by the presence of the others and the relative importance of electric and magnetic fields in the interaction effect varies among the tertiary circuits (except for the ideal case of "non-dissipative" circuits). In a practical case, therefore, the interaction coefficient is only approximately proportional to F_i. Measurements of the interaction coefficients by indirect methods have, however, indicated that the approximation is satisfactory for purposes of practical transposition design.

Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective

Basic Requirements*

By HARVEY FLETCHER

The fundamental requirements involved in a system capable of picking up orchestral music, transmitting it a long distance, and reproducing it in a large hall, are discussed in this paper.

IN THIS electrical era one is not surprised to hear that orchestral music can be picked up in one city, transmitted a long distance, and reproduced in another. Indeed, most people think such things are commonplace. They are heard every night on the radio. However, anyone who appreciates good music would not admit that listening even to the best radio gives the emotional thrill experienced in the concert hall. Nor is it evident that a listener in a small room ever will be able to get the same effect as that experienced in a large hall, although it must be admitted that such a question is debatable. The proper answer will involve more than a consideration of only the physical factors.

This symposium describes principles and apparatus involved in the reproduction of music in large halls, the reproduction being of a character that may give even greater emotional thrills to music lovers than those experienced from the original music. This statement is based upon the testimony of those who have heard some of the few concerts reproduced by the apparatus which will be described in the papers of this symposium.

It is well known that when an orchestra plays, vibrations which are continually changing in form are produced in the air of the concert hall where the orchestra is located. An ideal transmission and reproducing system may be considered as one that produces a similar set of vibrations in a distant concert hall in which is executed the same time-sequence of changes that takes place in the original hall. Since such changes are different at different positions in the hall, the use of such an ideal system implies that at corresponding positions in the

* First paper in the Symposium. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering*, January, 1934.

two halls this time-sequence should be the same. Obviously, this never can be true at every position unless the halls are the same size and shape; corresponding positions would not otherwise exist. Let us consider the case where the two halls are the same size and shape and also have the same acoustical properties. Let us designate the first hall in which the music originates by O, and the second one in which the music is reproduced by R. What requirements are necessary to obtain perfect reproduction from O into R such that any listener in any part of R will receive the same sound effects as if he were in the corresponding position in O?

Suppose there were interposed between the orchestra and the audience a flexible curtain of such a nature that it did not interfere with a free passage of the sound, and which at the same time had scattered uniformly over it microphones which would pick up the sound waves and produce a faithful electrical copy of them. Assume each microphone to be connected with a perfect transmission line which terminates in a projector occupying a corresponding position on a similar curtain in hall R. By a perfect transmission line is meant one that delivers to the projector electrical energy equal both in form and magnitude to that which it receives from the microphone. If these sound projectors faithfully transform the electrical vibrations into sound vibrations, the audience in hall R should obtain the same effect as those listening to the original music in hall O.

Theoretically, there should be an infinite number of such ideal sets of microphones and sound projectors, and each one should be infinitesimally small. Practically, however, when the audience is at a considerable distance from the orchestra, as usually is the case, only a few of these sets are needed to give good auditory perspective; that is, to give depth and a sense of extensiveness to the source of the music. The arrangement of some of these simple systems, together with their effect upon listeners in various parts of the hall, is described in the paper by Steinberg and Snow (page 245).

In any practical system it is important to know how near these ideal requirements one must approach before the listener will be aware that there has been any degradation from the ideal. For example, it is well known that whenever a sound is suddenly stopped or started, the frequency band required to transmit the change faithfully is infinitely wide. Theoretically, then, in order to fulfill these ideal requirements for transmitting such sounds, all three elements in the transmission system should transmit all possible frequencies without change. Practically, because of the limitations of hearing, this is not necessary. If the intensities of some of the component frequencies required to represent such a change are below the threshold of audibility it is obvious that their elimination will not be detected by the average normal ear. Consequently, for highgrade reproduction of sounds it is obvious that, except in very special cases, the range of frequencies that the system must transmit is determined by the range of hearing rather than by the kind of sound that is being reproduced.

Tests have indicated that, for those having normal hearing, pure tones ranging in frequency from 20 to 20,000 cycles per second can be heard. In order to sense the sounds at either of these extreme limits, they must have very high intensity. In music these frequencies usually are at such low intensities that the elimination of frequencies below 40 c.p.s. and those above 15,000 c.p.s. produces no detectable difference in the reproduction of symphonic music. These same tests also indicated that the further elimination of frequencies beyond either of these limits did begin to produce noticeable effects, particularly on a certain class of sounds produced in the orchestra. For example, the elimination of all frequencies above 13,000 c.p.s. produced a detectable change in the reproduced sound of the snare drum, cymbals, and castanets. Also, the elimination of frequencies below 40 c.p.s. produced detectable differences in reproduced music of the base viol, the bass tuba, and particularly of the organ.

Within this range of frequencies the system (the combination of the microphone, transmission line, and loud speaker) should reproduce the various frequencies with the same efficiency. Such a general statement sounds correct, but a careful analysis of it would reveal that when any one tried to build such a system or tried to meet such a requirement he would have great difficulty in understanding what it meant.

For example, for reproducing all the frequencies within this band, a certain system may be said to have a uniform efficiency when it operates between two rooms under the condition that the pressure variation at a certain distance away from the sound projector is the same as the pressure variation at a certain position in front of the microphone. It is obvious, however, that in other positions in the two rooms this relation would not in general hold. Also, if the system were transferred into another pair of rooms the situation would be entirely changed. These difficulties and the way they were met are discussed in the papers of this symposium that deal with loud speakers and microphones (p. 259) and with methods of applying the reproducing system to the concert hall (p. 301). It will be obvious from these papers that the criterion for determining the ideal frequency characteristics to be given to the system is arbitrary within certain limits. However, solving the problem according to criteria adopted produced a system that gave very satisfactory results.

Besides the requirement on frequency response just discussed, the system also must be capable of handling sound powers that vary through a very wide range. If this discussion were limited to the type of symphonic music that now is produced by the large orchestras, this range would be about 10,000,000 to 1, or 70 decibels. To reproduce such music then, the system should be capable of handling the smallest amount of power without introducing extraneous noises approaching it in intensity, and also reproduce the most intense sounds without overloading any part of the transmission system. However, this range is determined by the capacities of the musical instruments now available and the man power that conveniently can be grouped together under one conductor. As soon as a system was built that was capable of handling a much wider range, the musicians immediately took advantage of it to produce certain effects that they previously had tried to obtain with the orchestra alone, but without success because of the limited power of the instruments themselves. For these reasons it seems clear that the desirable requirements for intensity range, as well as those for frequency range, are determined largely by the ear rather than by the physical characteristic of any sound. An ideal transmission should, without introducing an extraneous audible sound, be capable of reproducing a sound as faintly as the ear can hear and as loudly as the ear can tolerate. Such a range has been determined with the average normal ear when using pure tones. The results of recent tests are shown in Fig. 1.

The ordinates are given in decibels above the reference intensity which is 10^{-16} watts per square centimeter. The values are for field intensities existing in an air space free from reflecting walls. The most intense peaks in music come in the range between 200 and 1000 c.p.s. Taking an average for this range it may be seen that there is approximately a 100-db range in intensity for the music, provided about 10 db is allowed for the masking of sound in the concert hall even when the audience is quietest.

The music from the largest orchestra utilizes only 70 db of this range when it plays in a concert hall of usual size. To utilize the full capabilities of the hearing range the ideal transmission system should add about 10 db on the *pp* side and 20 db on the *ff* side of the range. The capacity of the sound projectors necessary to reach the maximum allowable sound that the ear can tolerate varies with the size of the room. A good estimate can be obtained by the following consideration.

If T is the time of reverberation of the hall in seconds, E the power of the sound source in watts, I the maximum energy density per cubic centimeter in joules, and V the volume of the hall in cubic centimeters, then it is well known that

$$I = \frac{1}{6 \log_e 10} \cdot \frac{ET}{V}. \tag{1}$$

Measurements have shown that when the sound intensity in a free field reaches about 10^{-4} watts per square centimeter, the average person begins to *feel* the sound. This maximum value is approxi-

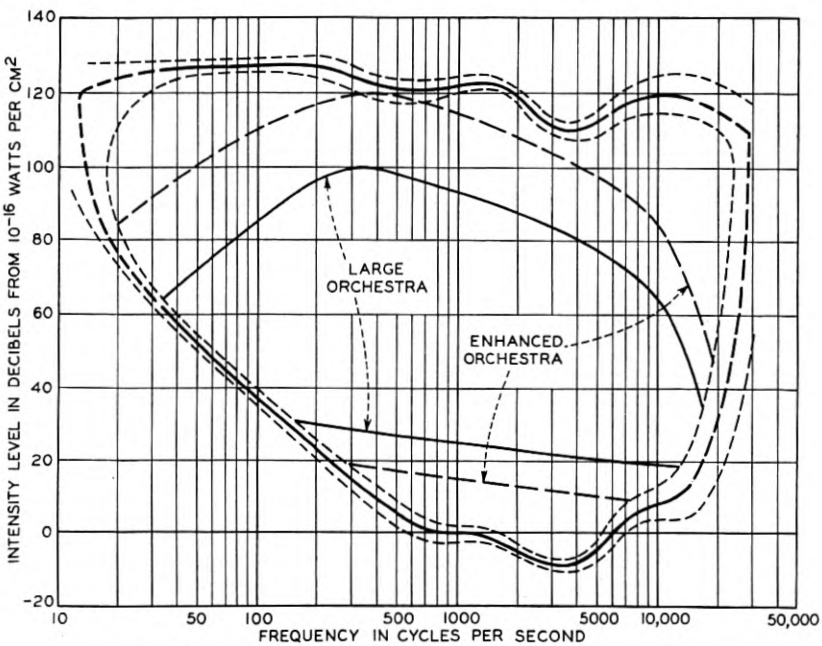

Fig. 1—Limits of audible sound as determined by recent tests.

mately the same for all frequencies in the important audible range. Any higher intensities, and for some persons somewhat lower intensities, become painful and may injure the hearing mechanism. This intensity corresponds to an energy density I of 3×10^{-9} joules. Using this figure as the upper limit to be tolerated by the human ear, then, the maximum power of the sound source must be given by

$$E = 4.1 \times 10^{-8} \frac{V}{T}. \tag{2}$$

For halls like the Academy of Music in Philadelphia and Carnegie Hall in New York City, in which the volume V is approximately 2×10^{10} cubic centimeters and the reverberation time about 2 seconds,

E, the power of the sound source, is approximately 400 watts. For other halls it may be seen that the power required for this source is proportional to the volume of the hall and inversely proportional to the reverberation time. A person would experience the sense of *feeling* when closer than about 10 meters to such a source of 400 watts power, even in free open space. Hence it would be unwise to have seats closer than 10 or 15 meters from the stage when such powers are to be used.

These, then, are the general fundamental requirements for an ideal transmission system. How near they can be realized with apparatus that we now know how to build will be discussed in the papers included in this symposium.

A system approximately fulfilling these requirements was constructed and used to reproduce the music played by the Philadelphia Orchestra. The first public demonstration was given in Constitution Hall, Washington, D. C., on the evening of April 27, 1933, under the auspices of the National Academy of Sciences. At that time, Dr. Stokowski, Director of the Philadelphia Orchestra, manipulated the electric controls from a box in the rear of Constitution Hall while the orchestra, led by Associate Conductor Smallens, played in the Academy of Music in Philadelphia.

Three microphones of the type described in the paper by Wente and Thuras (p. 259) were placed before the orchestra in Philadelphia, one on each side and one in the center at about 20 feet in front of and 10 feet above the first row of instruments in the orchestra. The electrical vibrations generated in each of these microphones were amplied by voltage amplifiers and then fed into a transmission line which was extended to Washington by means of telephone cable. The construction of these lines, the equipment used with them, and their electrical properties, are described in the paper by Affel, Chesnut, and Mills (p. 285). In Constitution Hall at Washington, D. C., these transmission lines were connected to power amplifiers. The type of power amplifiers and voltage amplifiers used are described in the paper by Scriven (p. 278). The output of these amplifiers fed three sets of loud speakers like those described in the paper by Wente and Thuras. They were placed on the stage in Constitution Hall in positions corresponding to the microphones in the Academy of Music, Philadelphia.

Judging from the expression of those who heard this concert, the development of this system has opened many new possibilities for the reproduction and transmission of music that will create even a greater emotional appeal than that obtained when listening to the music coming directly from the orchestra through the air.

Physical Factors*

By J. C. STEINBERG and W. B. SNOW

In considering the physical factors affecting it, auditory perspective is defined in this paper as being reproduction which preserves the spatial relationships of the original sounds. Ideally, this would require an infinite number of separate microphone-to-speaker channels; practically, it is shown that good auditory perspective can be obtained with only 2 or 3 channels.

ABILITY to localize the direction, and to form some judgment of the distance from a sound source under ordinary conditions of listening, are matters of common experience. Because of this faculty an audience, when listening directly to an orchestral production, senses the spatial relations of the instruments of the orchestra. This spatial character of the sounds gives to the music a sense of depth and of extensiveness, and for perfect reproduction should be preserved. In other words, the sounds should be reproduced in true *auditory perspective.*

In the ordinary methods of reproduction, where only a single loud speaking system is used, the spatial character of the original sound is imperfectly preserved. Some of the depth properties of the original sound may be conveyed by such a system,[1] but the directional properties are lost because the audience tends to localize the sound as coming from the direction of a single source, the loud speaker. Ideally, there are two ways of reproducing sounds in true auditory perspective. One is binaural reproduction which aims to reproduce in a distant listener's ears, by means of head receivers, exact copies of the sound vibrations that would exist in his ears if he were listening directly. The other method, which was described in the first paper of this series, uses loud speakers and aims to reproduce in a distant hall an exact copy of the pattern of sound vibration that exists in the original hall. In the limit, an infinite number of microphones and loud speakers of infinitesimal dimensions would be needed.

Far less ideal arrangements, consisting of as few as two microphone-loudspeaker sets, have been found to give good auditory perspective. Hence, it is not necessary to reproduce in the distant hall an exact copy of the vibrations existing in the original hall. What physical

* Second paper in the Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering,* January, 1934.

properties of the waves must be preserved then, and how are these properties preserved by various arrangements of 2- and 3-channel loudspeaker reproducing systems? To answer these questions, some very simple localization tests have been made with such systems. Perhaps attention can be focused more easily on their important properties by considering briefly the results of these tests.

Localization Afforded by Multichannel Systems

In Fig. 1 is shown a diagram of the experimental set-up that was used. The microphones, designated as *LM* (left), *CM* (center), and *RM* (right), were set on a "pick-up" stage that was marked out on the floor of an acoustically treated room. The loud speakers, designated as *LS*, *CS*, and *RS*, were placed in the front end of the auditorium at the Bell Telephone Laboratories and were concealed from view by a curtain of theatrical gauze. The average position of a group of twelve observers is indicated by the cross in the rear center part of the auditorium.

The object of the tests was to determine how a caller's position on the pick-up stage compared with his apparent position as judged by the group of observers in the auditorium listening to the reproduced speech. Words were uttered from some 15 positions on the pick-up stage in random order. The 9 positions shown in Fig. 1 were always included in the 15, the remaining positions being introduced to minimize memory effects. The reproducing system was switched off while the caller moved from one position to the other.

In the first series of tests, the majority of the observers had no previous experience with the set-up. They simply were given a sheet of coordinate paper with a single line ruled on it to indicate the line of the gauze curtain and asked to locate the apparent position of the caller with respect to this line. Following these tests, the observers were permitted to listen to speech from various announced positions on the pick-up stage. This gave them some notion of the approximate outline of what might be called the "virtual" stage. These tests then were repeated. As there was no significant difference in results, the data from both tests have been averaged and are shown in Fig. 1.

The small diagram at the top of Fig. 1 shows the caller's positions with respect to the microphone positions on the pick-up stage. The corresponding average apparent positions when reproduced are shown with respect to the curtain line and the loudspeaker positions. The type of reproduction is indicated symbolically to the right of the apparent position diagrams.

With 3-channel reproduction there is a reasonably good corre-

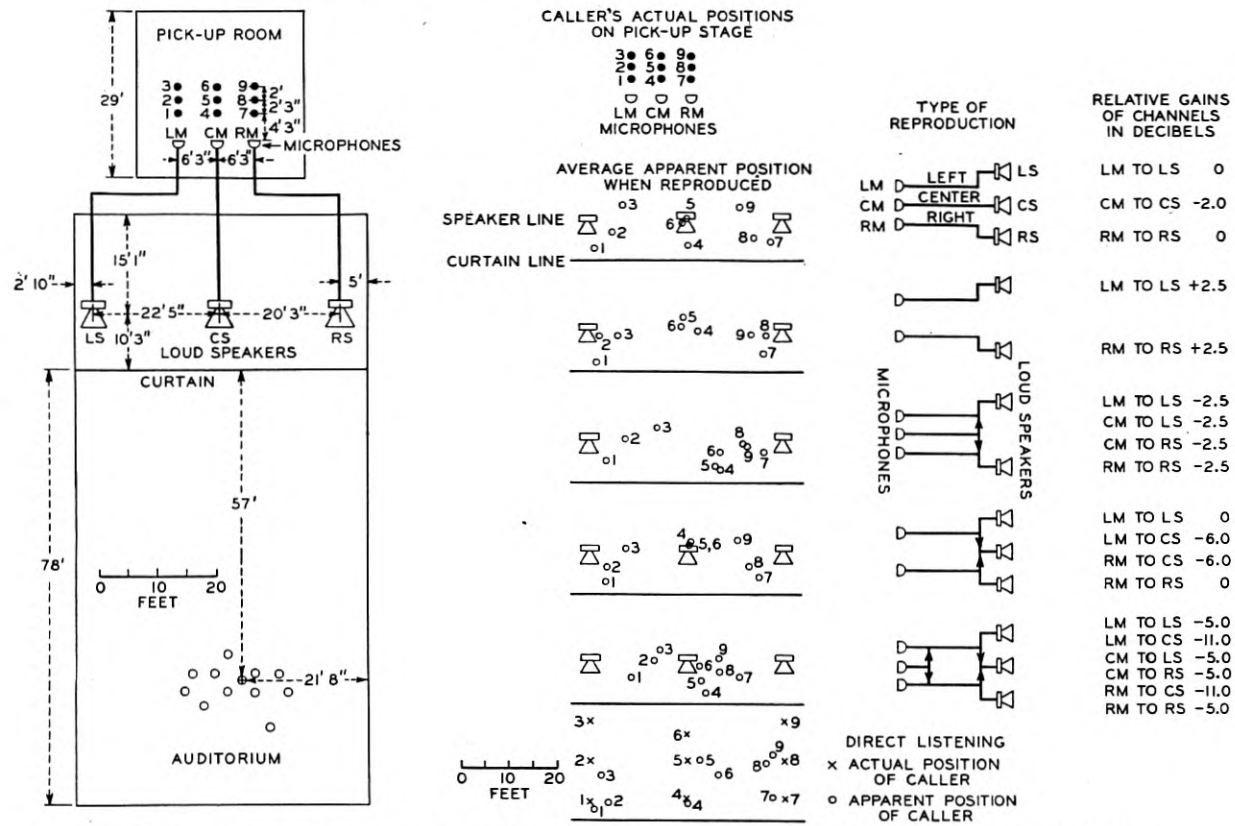

Fig. 1—Diagram of arrangement (left) for sound localization tests and (right) the results obtained.

spondence between the caller's actual position on the pick-up stage and his apparent position on the virtual stage. Apparent positions to the right or left correspond with actual positions to the right or left, and apparent front and rear positions correspond with actual front and rear positions. Thus the system afforded lateral or "angular" localization as well as fore and aft or "depth" localization. For comparison, there is shown in the last diagram the localization afforded by direct listening. The crosses indicate a caller's position in back of the gauze curtain and the circles indicate his apparent position as judged by the observers listening to his speech directly. In both cases, as the caller moved back in a straight line on the left or right side of the stage, he appeared to follow a curved path pulling in toward the rear center; e.g., compare the caller positions 1, 2, 3, with the apparent positions 1, 2, 3. This distortion was somewhat greater for 3-channel reproduction than for direct listening.

The results obtained with the 2-channel system show two marked differences from those obtained with 3-channel reproduction. Positions on the center line of the pick-up stage (i.e., 4, 5, 6) all appear in the rear center of the virtual stage, and the virtual stage depth for all positions is reduced. The virtual stage width, however, is somewhat greater than that obtained with 3-channel reproduction.

Bridging a third microphone across the 2-channel system had the effect of pulling the center line positions 4, 5, 6, forward, but the virtual stage depth remained substantially that afforded by 2-channel reproduction, while the virtual stage width was decreased somewhat. In this and the other bridged arrangements the bridging circuits employed amplifiers, as represented by the arrows in Fig. 1, in such a way that there was a path for speech current only in the indicated direction.

Bridging a third loud speaker across the 2-channel system had the effect of increasing the virtual stage depth and decreasing the virtual stage width, but positions on the center line of the pick-up stage appeared in the rear center of the virtual stage as in 2-channel reproduction.

Bridging both a third microphone and a third loud speaker across the 2-channel system had the effect of reducing greatly the virtual stage width. The width could be restored by reducing the bridging gains, but fading the bridged microphone out caused the front line of the virtual stage to recede at the center, whereas fading the bridged loud speaker out reduced the virtual stage depth. No fixed set of bridging gains was found that would enable the arrangement to create the virtual stage created by three independent channels. The gains used in obtaining the data shown in Fig. 1 are indicated at the right of the symbolic circuit diagrams.

Factors Affecting Depth Localization

Before attempting to explain the results that have been given in the foregoing, it may be of interest to consider certain additional observations that bear more specifically upon the factors that enter into the "depth" and "angular" localization of sounds. The microphones on the pick-up stage receive both direct and reverberant sound, the latter being sound waves that have been reflected about the room in which the pick-up stage is located. Similarly, the observer receives the reproduced sounds directly and also as reverberant sound caused by reflections about the room in which he listens. To determine the effects of these factors, the following three tests were made:

1. Caller remained stationary on the pick-up stage and close to microphone, but the loudness of the sound received by the observer was reduced by gain control. This was loudness change without a change in ratio of direct to reverberant sound intensity.

2. Caller moved back from microphone, but gain was increased to keep constant the loudness of the sound received by the observer. This was a change in the ratio of direct to reverberant sound intensity without a loudness change.

3. Caller moved back from microphone, but no changes were made in the gain of the reproducing system. This changed both the ratio and the loudness.

All of the observers agreed that the caller appeared definitely to recede in all three cases. That is, either a reduction in loudness or a decrease in ratio of direct to reverberant sound intensity, or both, caused the sound to appear to move away from the observer. Position tests using variable reverberation with a given pick-up stage outline showed that increasing the reverberation moved the front line of the virtual stage toward the rear, but had slight effect upon the rear line. When the microphones were placed outdoors to eliminate reverberation, reducing the loudness either by changing circuit gains or by increasing the distance between caller and microphone moved the whole virtual stage farther away. It is because of these effects that all center line positions on the pick-up stage appeared at the rear of the virtual stage for 2-channel reproduction.

It has not been found possible to put these relationships on a quantitative basis. Probably a given loudness change, or a given change in ratio of direct to reverberant sound intensity, causes different sensations of depth depending upon the character of the reproduced sound and upon the observer's familiarity with the acoustic conditions surrounding the reproduction. Since the depth localization is inaccurate even when listening directly, it is difficult to obtain sufficiently accurate data to be of much use in a quantitative way. Because of this inaccuracy, good auditory perspective may be obtained with reproduced sounds even though the properties controlling depth localization depart materially from those of the original sound.

Angular Localization

Fortunately, the properties entering into lateral or angular localization permit more quantitative treatment. In dealing with angular localization, it has been found convenient to neglect entirely the effects of reverberant sound and to deal only with the properties of the sound waves reaching the observer's ears without reflections. The reflected waves or reverberant sounds do appear to have a small effect on angular localization, but it has not been found possible to deal with such sound in a quantitative way. One of the difficulties is that, because of differences in the build-up times of the direct and reflected sound waves, the amount of direct sound relative to reverberant sound reaching the observer's ears for impulsive sounds such as speech and music is much greater than would be expected from steady state methods of dealing with reverberant sound.

For the case of a plane progressive wave from a single sound source, and where the observer's head is held in a fixed position, there are apparently only three factors that can assist in angular localization: namely, phase difference, loudness difference, and quality difference between the sounds received by the two ears.

In applying these factors to the localization of sounds from more than one source, as in the present case, the effects of phase differences have been neglected. It is difficult to see how phase differences in this case can assist in localization in the ordinary way. The two remaining factors, loudness and quality differences, both arise from the directivity of hearing. This directivity probably is due in part to the shadow and diffraction effects of the head and to the differences in the angle subtended by the ear openings. Measurements of the directivity with a source of pure tone located in various positions around the head in a horizontal plane have been reported by Sivian and White.[2] From these measurements, the loudness level differences between near and far ears have been determined for various frequencies. These differences are shown in Fig. 2 from which, using the pure tone data given, similar loudness level differences for complex tones may be calculated. Such calculated differences for speech are shown in Fig. 3.

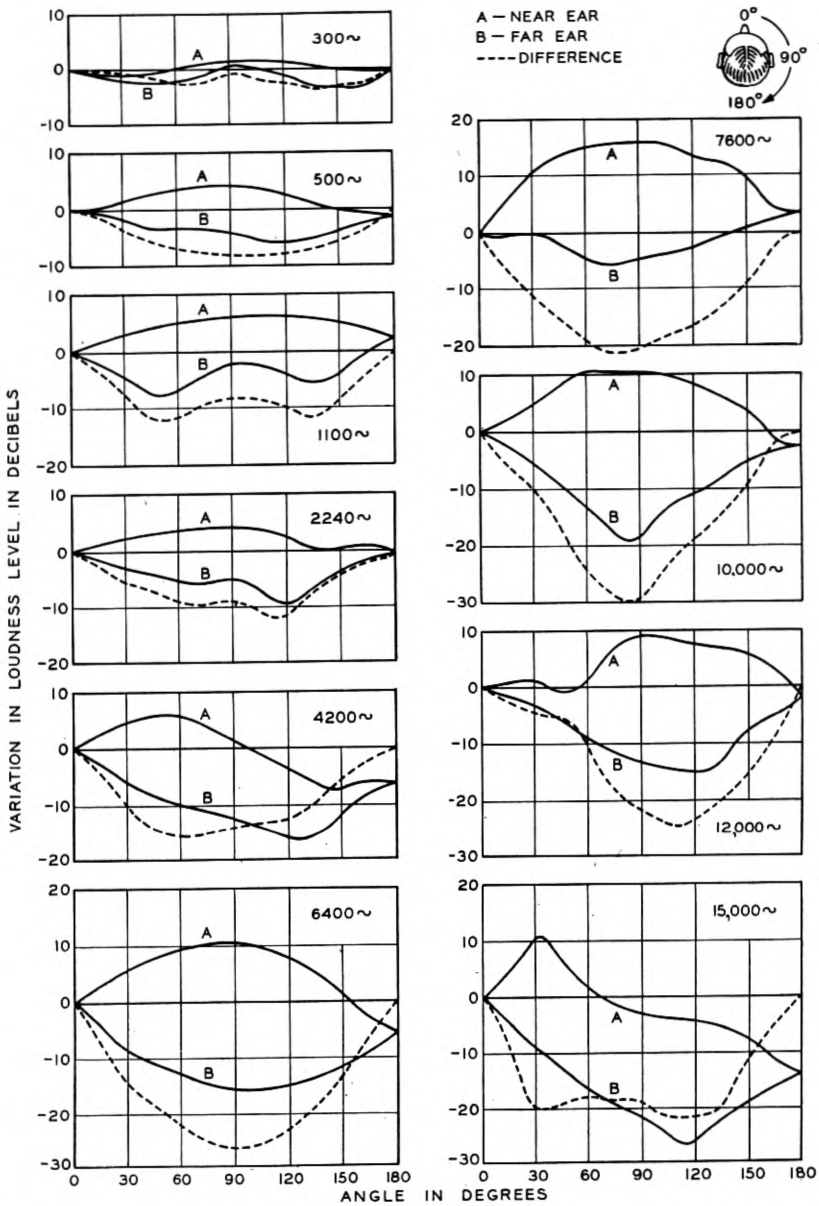

Fig. 2—Variation in loudness level as a sound source is rotated in a horizontal plane around the head.

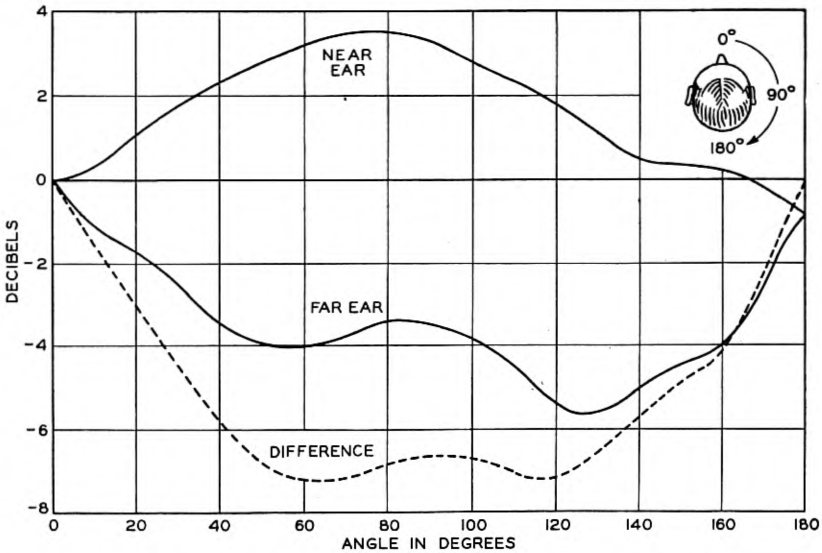

Fig. 3—Variation in loudness as a speech source is rotated in a horizontal plane around the head.

As may be inferred from the varying shapes of the curves of Fig. 2, the directive effects of hearing introduce a frequency distortion more or less characteristic of the direction from which the sound comes. Thus the character or quality of complex sounds varies with the angle of the source. There are quality differences at each ear for various angles of source, and quality differences between the two ears for a given angle of source. In Fig. 4 is shown the frequency distortion at the right ear when a source of sound is moved from a position on the right to one on the left of an observer. It is a graph of the "difference" values of Fig. 2 for an angle of 90 degrees. Frequencies above 4,000 cycles per second are reduced by as much as 15 to 30 decibels. This amount of distortion is sufficient to affect materially the quality of speech, particularly as regards the loudness of the sibilant sounds.

Reference to the difference curve of Fig. 3 shows that if, for example, a source of speech is 20 degrees to the right of the median plane the speech heard by the right ear is 3 db louder than that heard by the left ear. A similar difference exists when the angle is 167 degrees. Presumably, when the right ear hears speech 3 db louder than the left, the observer localizes the sound as coming from a position 20 degrees or 167 degrees to the right, depending upon the quality of the speech. If this be assumed to be true, even though the difference is caused by the combination of sounds of similar quality from several sources, it should be possible to calculate the apparent angle.

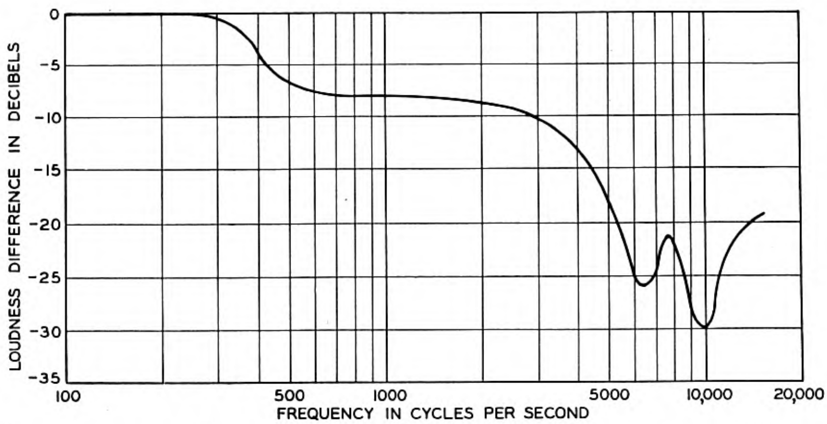

Fig. 4—Loudness difference produced in the right ear when a source of pure tone is moved from the right to the left of an observer.

Loudness Theory of Localization

Upon this assumption the apparent angle of the source as a function of the difference in decibels between the speech levels emitted by the loud speakers of the 2- and 3-channel systems has been calculated. Each loud speaker contributes an amount of direct sound loudness to each ear, depending upon its distance from, and its angular position with respect to, the observer. These contributions were combined on a power basis to give a resultant loudness of direct sound at each ear, from which the difference in loudness between the two ears was determined. The calculated results for the 2- and 3-channel systems are shown by the solid lines in Fig. 5. The y axis shows the apparent angle, positive angle being measured in a clockwise direction. The x axis shows the difference in decibels between the speech levels from the right and left loud speakers. The points are observed values taken from Fig. 1. The observed apparent angles were obtained directly from the average observer's location and the average apparent positions shown in Fig. 1. The speech levels from each of the loud speakers were calculated for each position on the pick-up stage. This was done by assuming that the waves arriving at the microphone had relative levels inversely proportional to the squares of the distances traversed. By correcting for the angle of incidence and for the known relative gains of the systems, the speech levels from the loud speakers were obtained.

A comparison of the observed and calculated results seems to indicate that the loudness difference at the two ears accounts for the greater part of the apparent angle of the reproduced sounds. If this is true, the angular location of each position on the virtual stage results from a particular loudness difference at the two ears produced by the speech coming from the loud speakers. When three channels are used a definite

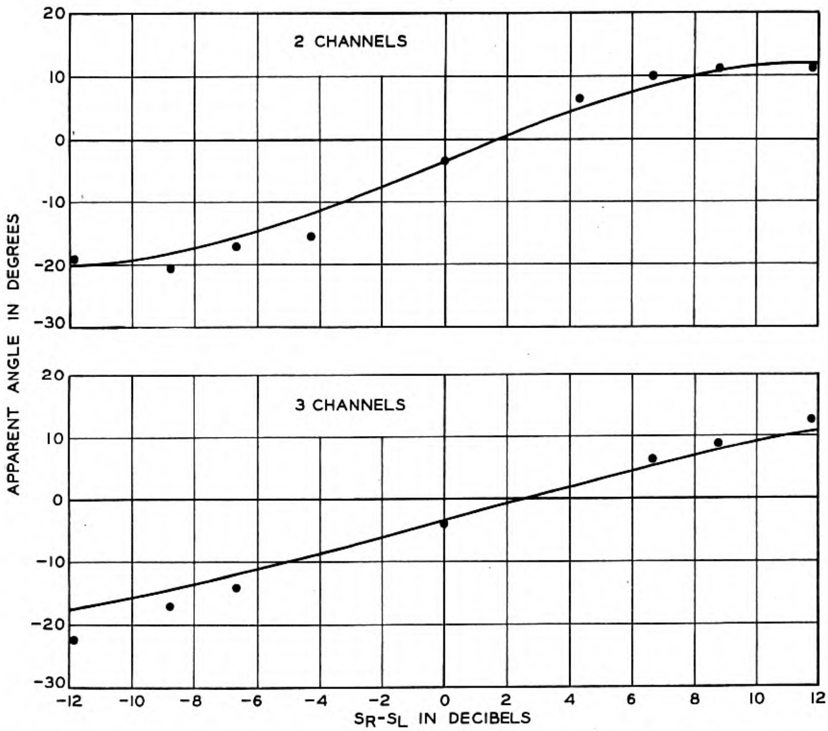

Fig. 5—Calculated and observed apparent angles for 2- and 3-channel reproduction.

set of loud speaker speech levels exists for each position on the pick-up stage. To create these same sets of loud speaker speech levels with the 3-microphone 3-loud speaker bridging arrangement already discussed, it would be necessary to change the bridging gains for each position on the pick-up stage. Hence it could not be expected that the arrangement as used (i.e., with fixed gains) would create a virtual stage identical with that created by 3-channel reproduction. However, with proper technique, bridging arrangements on a given number of channels can be made to give better reproduction than would be obtained with the channels alone.

Experimental Verification of Theory

Considerations of loudness difference indicate that all caller positions on the pick-up stage giving the same relative loud speaker outputs should be localized at the same virtual angle. The solid lines of Fig. 6 show a stage layout used to test this hypothesis with the 2-channel system. All points on each line have a constant ratio of distances to

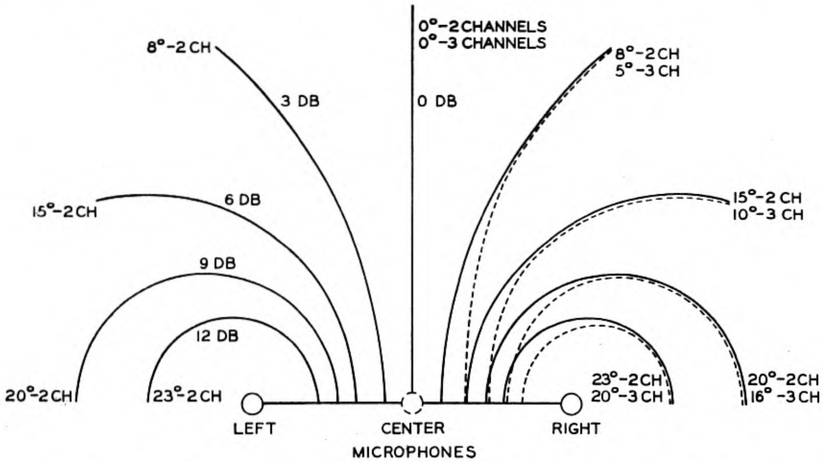

Fig. 6—Pick-up stage contour lines of constant apparent angle.

the microphones. The resulting direct sound differences in pressure expressed in decibels and the corresponding calculated apparent angles are indicated beside the curves. The apparent angles were calculated for an observing position on a line midway between the two loud speakers but at a distance from them equal to the separation between them. The microphones were turned face up at the height of the talker's lips to eliminate quality changes caused by changing incidence angle. It was found that a caller walking along one of these lines maintained a fairly constant virtual angle. For caller positions far from the microphones the observed angles were somewhat greater than those computed. For highly reverberant conditions, the tendency was toward greater calculated than observed angles. Reverberation also decreased the accuracy of localization.

A change of relative channel gain caused a change in virtual angle as would be expected from loudness difference considerations. For instance, if the caller actually walked the left 3-db line, he seemed to be on the 6-db line when the left channel gain was raised 3 db. Many of the effects of moving about the pick-up stage could be duplicated by volume control manipulation as the caller walked forward and backward on the center path. With a bridged center microphone substituted for the two side microphones similar effects were possible and, in addition, the caller by speaking close to the microphone could be brought to the front of the virtual stage.

For observing positions near the center of the auditorium the observed angles agreed reasonably well with calculations based only upon loudness differences. As the observer moved to one side, however, the virtual source shifted more rapidly toward the nearer loud speaker than was predicted by the computations. This was true of reproduction in the auditorium, both empty and with damping simulating an audience, and outdoors on the roof. Computations and experiment also show a change in apparent angle as the observer moves from front to rear, but its magnitude is smaller than the error of an individual localization observation. Consequently, observers in different parts of the auditorium localize given points on the pick-up stage at different virtual angles.

Because the levels at the three microphones are not independent, and because the desired contours depend upon the effects at the ears, a 3-channel stage is not as simple to lay out as a 2-channel stage. For a given observing position, however, a set of contour lines can be calculated. The dashed lines at the right of Fig. 6 show four contours thus calculated for the circuit condition of Fig. 1 and the observing position previously mentioned. The addition of the center channel reduces the virtual angle for any given position on the pick-up stage by reducing the resultant loudness difference at the ears. Although the 3-channel contours approach the 2-channel contours in shape at the back of the stage, a given contour results in a greater virtual angle for 2- than for 3-channel reproduction.

Similar effects were obtained experimentally. As in 2-channel reproduction, movements of the caller could be simulated by manipulation of the channel gains. From an observing standpoint the 3-channel system was found to have an important advantage over the 2-channel system in that the shift of the virtual position for side observing positions was smaller.

Effects of Quality

If the quality from the various loud speakers differs, the quality of sound is important to localization. When the 2-channel microphones were so arranged that one picked up direct sound and reverberation while the other picked up mostly reverberation, the virtual source was localized exactly in the "direct" loud speaker until the power from the "reverberant" loud speaker was from 8 to 10 db greater. In general, localization tends toward the channel giving most natural or "closeup" reproduction, and this effect can be used to aid the loudness differences in producing angular localization.

Principal Conclusions

The principal conclusions that have been drawn from these investigations may be summarized as follows:

1. Of the factors influencing angular localization, loudness difference of direct sound seems to play the most important part; for certain observing positions the effects can be predicted reasonably well from computations. When large quality differences exist between the loudspeaker outputs, the localization tends toward the more natural source. Reverberation appears to be of minor importance unless excessive.

2. Depth localization was found to vary with changes in loudness, the ratio of direct to reverberant sound, or both, and in a manner not found subject to computational treatment. The actual ratio of direct to reverberant sound, and the change in the ratio, both appeared to play a part in an observer's judgment of stage depth.

3. Observers in various parts of the auditorium localize a given source at different virtual positions, as is predicted by loudness computations. The virtual source shifts to the side of the stage as the observer moves toward the side of the auditorium. Although quantitative data have not been obtained, qualitative data on these effects indicate that the observed shift is considerably greater than that computed. Moving backward and forward in the auditorium appears to have only a small effect on the virtual position.

4. Because of these physical factors controlling auditory perspective, point-for-point correlation between pick-up stage and virtual stage positions is not obtained for 2- and 3-channel systems. However, with stage shapes based upon the ideas of Fig. 7, and with suitable use of quality and reverberation, good auditory perspective can be produced. Manipulation of circuit conditions probably can be used advantageously to heighten the illusions or to produce novel effects.

5. The 3-channel system proved definitely superior to the 2-channel by eliminating the recession of the center-stage positions and in reducing the differences in localization for various observing positions. For musical reproduction, the center channel can be used for independent control of soloist renditions. Although the bridged systems did not duplicate the performance of the physical third channel, it is believed that with suitably developed technique their use will improve 2-channel reproduction in many cases.

6. The application of acoustic perspective to orchestral reproduction in large auditoriums gives more satisfactory performance than probably would be suggested by the foregoing discussions. The instruments near the front are localized by every one near their correct positions.

In the ordinary orchestral arrangement, the rear instruments will be displaced in the reproduction depending upon the listener's position, but the important aspect is that every auditor hears differing sounds from differing places on the stage and is not particularly critical of the exact apparent positions of the sounds so long as he receives a spatial impression. Consequently 2-channel reproduction of orchestral music gives good satisfaction, and the difference between it and 3-channel reproduction for music probably is less than for speech reproduction or the reproduction of sounds from moving sources.

References

1. "Some Physical Factors Affecting the Illusion in Sound Motion Pictures," J. P. Maxfield. *Jour. Acous. Soc.*, July, 1931.
2. "Minimum Audible Sound Fields," L. J. Sivian and S. D. White. *Jour. Acous. Soc.*, April, 1933.

Loud Speakers and Microphones*

By E. C. WENTE and A. L. THURAS

In ordinary radio broadcast of symphony music, the effort is to create the effect of taking the listener to the scene of the program, whereas in reproducing such music in a large hall before a large gathering the effect required is that of transporting the distant orchestra to the listeners. Lacking the visual diversion of watching the orchestra play, such an audience centers its interest more acutely in the music itself, thus requiring a high degree of perfection in the reproducing apparatus both as to quality and as to the illusion of localization of the various instruments. Principles of design of the loud speakers and microphones used in the Philadelphia-Washington experiment are treated at length in this paper.

AS EARLY as 1881 a large scale musical performance was reproduced by telephone instruments at the Paris Electrical Exhibition. Microphones were placed on the stage of the Grand Opera and connected by wires to head receivers at the exposition. It is interesting to note that separate channels were provided for each ear so as to give to the music perceived by the listener the "character of relief and localization." With head receivers it is necessary to generate enough sound of audible intensity to fill only a volume of space enclosed between the head receiver and the ear. As no amplifiers were available, the production of enough sound to fill a large auditorium would have been entirely outside the range of possibilities. With the advent of telephone amplifiers, microphone efficiency could be sacrificed to the interest of good quality where, as in the reproduction of music, this was of primary interest. When amplifiers of greater output power capacity were developed, loud speakers were introduced to convert a large part of the electrical power into sound so that it could be heard by an audience in a large auditorium. Improvements have been made in both microphones and loud speakers, resulting in very acceptable quality of reproduction of speech and music; as is found, for instance, in the better class of motion picture theaters.

In the reproduction, in a large hall, of the music of a symphony orchestra the approach to perfection that is needed to satisfy the habitual concert audience undoubtedly is closer than that demanded for any other type of musical performance. The interest of the listener here lies solely in the music. The reproduction therefore should be

* Third paper in the Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering*, January, 1934.

such as to give to a lover of symphonic music esthetic satisfaction at least as great as that which would be given by the orchestra itself playing in the same hall. This is more than a problem of instrument design, but this paper will be restricted to a discussion of the requirements that must be met by the loud speakers and microphones, and to a description of the principles of design of the instruments used in the transmission of the music of the Philadelphia Orchestra from Philadelphia to Constitution Hall in Washington. Some of the requirements are found in the results of measurements that have been made on the volume and frequency ranges of the music produced by the orchestra.

General Considerations

The acoustic powers delivered by the several instruments of a symphony orchestra, as well as by the orchestra as a whole, have been investigated by Sivian, Dunn, and White. Figure 1 was drawn on the basis of the values published by them.[1] The ordinates of the horizontal lines give the values of the peak powers within the octaves indicated by the positions of the lines. For a more exact interpretation of these values the reader is referred to the original paper, but the chart here given will serve to indicate the power that a loud speaker must be capable of delivering in the various frequency regions, if the reproduced music is to be as loud as that given by the orchestra itself. However, it was the plan in the Philadelphia-Washington experiment to reproduce the orchestra, when desired, at a level 8 or 10 db higher, so that with three channels each loud speaking system had to be able to deliver two or three times the powers indicated in Fig. 1. Sivian, Dunn, and White also found that for the whole frequency band the peak powers in some cases reached values as high as 65 watts. In order to go 8 db above this value, each channel would have to be capable of delivering in the neighborhood of 135 watts.

The chart (Fig. 1) shows that the orchestra delivers sound of comparable intensity throughout practically the whole audible range. Although it is conceivable that the ear would not be capable of detecting a change in quality if some of the higher or lower frequencies were suppressed, measurements published by W. B. Snow [2] show that for any change in quality in any of the instruments to be undetectable the frequency band should extend from about 40 to about 13,000 c.p.s. The necessary frequency ranges that must be transmitted to obviate noticeable change in quality for the different orchestral instruments are indicated in the chart of Fig. 2, which is taken from the paper by Snow.

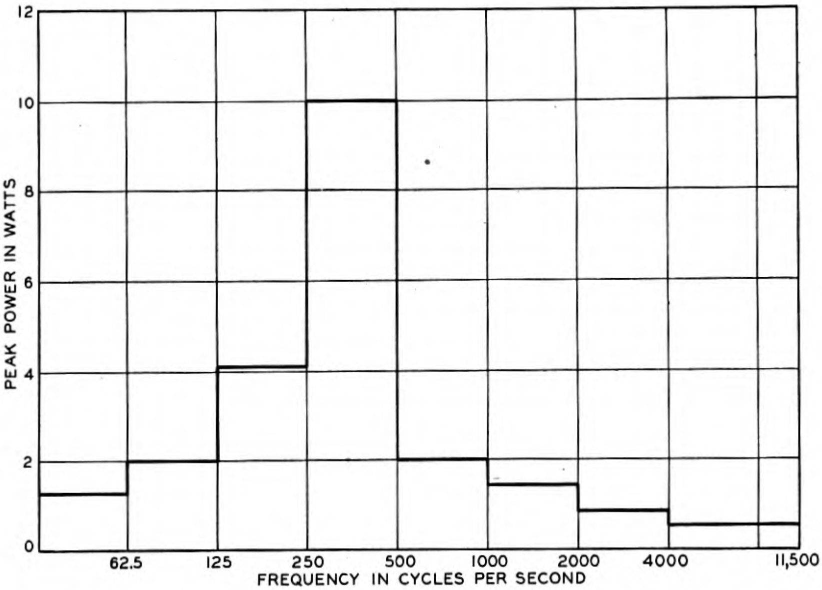

Fig. 1—Peak powers delivered by an orchestra within various frequency regions.

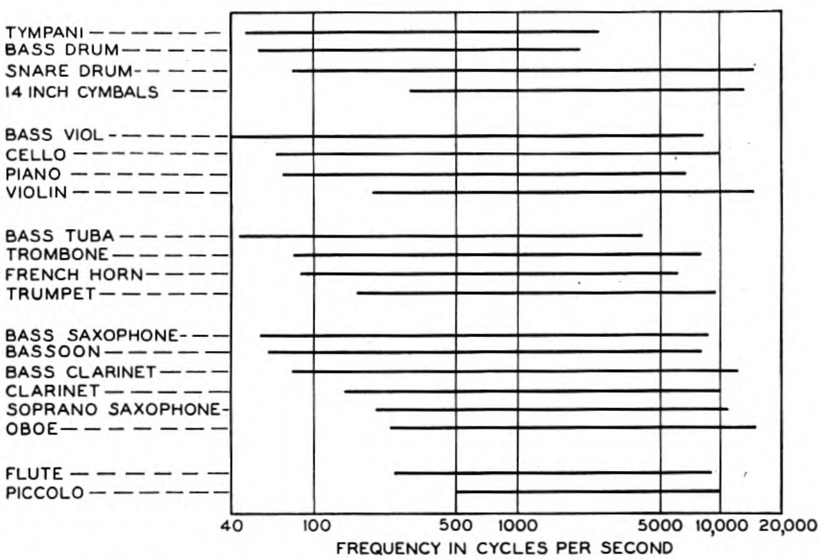

Fig. 2—Frequency transmission range required to produce no noticeable distortion for orchestral instruments.

Thus far only the sound generated by the orchestra itself has been considered. However, it is well known that the esthetic value of orchestral music in a concert hall is dependent to a very great extent upon the acoustic properties of the hall. At first thought one might be inclined to leave this out of account in considering the reproduction by a loud speaking system, as one should normally choose a hall known to have satisfactory acoustics for an actual orchestra. There would be no further problem in this if the orchestral instruments and the loud speaker radiated the sound uniformly in all directions, but some of the important instruments are quite directive; i.e., they radiate much the greater portion of their sound through a relatively small angle. As an example, a polar diagram giving the relative intensities of the sound radiated in various directions by the violin is given in Fig. 3, which is taken from a paper published by Backhaus.[3] The

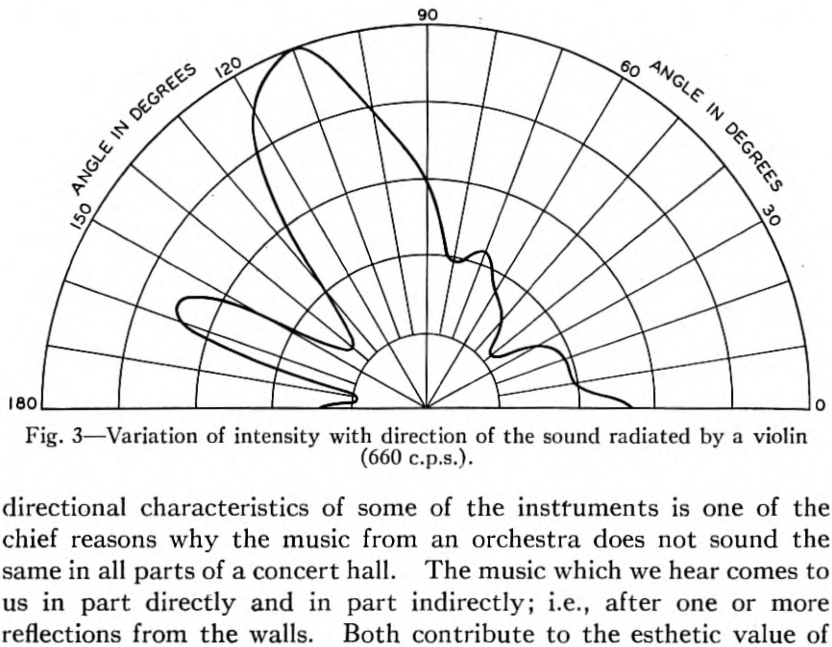

Fig. 3—Variation of intensity with direction of the sound radiated by a violin (660 c.p.s.).

directional characteristics of some of the instruments is one of the chief reasons why the music from an orchestra does not sound the same in all parts of a concert hall. The music which we hear comes to us in part directly and in part indirectly; i.e., after one or more reflections from the walls. Both contribute to the esthetic value of the music. The ratio of the direct to the indirect sound, which has been designated by Hughes[4] as the *acoustic ratio*, is to a first approximation inversely proportional to the product of the reverberation time and the angle through which the sound is radiated.[5] For a steady tone by far the greater part of the intensity at a given point in a hall remote from the source is attributable to the indirect sound. However, inasmuch as many of the tones of a musical selection are of short duration, the direct sound is of great importance; it is this sound alone which enables us to localize the source. So far as this ratio is concerned, a decrease in the radiating angle of a loud speaker is equivalent to a reduction in the reverberation time of the hall. The effect on the music, however, is not entirely equivalent, for the rate of decay of sound in the room is unaltered by a change in directivity of the source, as this depends only on the reverberation time.

As already pointed out, some of the instruments of the orchestra are quite directive and others are nondirectional. In general, it may be said that the instruments of lower register are less directive than those of higher register. To have each instrument as reproduced by the loud speaker sound just as the instrument itself would sound in the same hall, the loud speaker would have to reproduce the music from each instrument with a directivity corresponding to that of the instrument itself. This manifestly is impossible. The best that can be hoped for is a compromise. Let the loud speaking system be designed so that it is nondirective for the lower frequencies, and at the higher frequencies it will radiate the sound through a larger angle than the most directive of the instruments and through a smaller angle than the least directive. Although this compromise means that the individual instruments will not sound exactly like the originals, it carries with it one advantage: At all the seats in the hall included in the radiating angle and at a given distance from the loud speaker the music may be heard to equal advantage, whereas with the orchestra itself the most desirable seats comprise only a certain portion of the hall. The optimum radiating angle is largely a matter of judgment; if it is too small the music will lack the spatial quality experienced at indoor concerts; if it is too large there will be a loss in definition.

There is another respect in which the directivity of the source can greatly affect the tone quality. Most loud speakers radiate tones of low frequency through a relatively large angle, but as the frequency is increased this angle becomes smaller and smaller. Under this condition the relation between the intensities of the high and low frequency tones as received directly will be different for almost all parts of the hall. Hence, even with equalization by electrical networks, the reproduction at best can be good only at a few places in the hall. Therefore, the sound radiated not only should be contained within a certain solid angle, but the radiation throughout this angle should be uniform at all frequencies.

The Loud Speaker

At present two kinds of loud speakers are in wide commercial use, the direct radiating and the horn types. Each has its merits, but the latter was used in the Philadelphia-Washington experiment because it appears to have definite advantages where such large amounts of power are to be radiated. The horn type can be given the desired directive properties more readily, and higher values of efficiency throughout a wide frequency range are more easily realized. In consideration of the large power requirements, high efficiency is of special importance because it will keep to the lowest possible value the power capacity requirements of the amplifiers and because, with the heating proportional to one minus the efficiency, the danger of burning out the receiving units is reduced.

For efficiently radiating frequencies as low as 40 c.p.s., a horn of large dimensions is required. In order that the apparatus may not become too unwieldly the folded type of horn is preferable, but a large folded horn transmits high frequency tones very inefficiently. As actually used, therefore, the loud speaker was constructed in two units: one for the lower and the other for the higher frequencies, an electrical network being used to divide the current into two frequency bands, the point of division being about 300 c.p.s.

The Low Frequency Horn

When moderate amounts of power are transmitted through a horn the sound waves will suffer very little distortion, but when the power per unit area becomes large, second-order effects, usually neglected in considering waves of small amplitude, must be taken into account. The transmission of waves of large amplitude through an exponential horn has been investigated theoretically by M. Y. Rocard.[6] His investigation shows that if W watts are transmitted through the throat of an exponential horn a second harmonic of intensity RW will be generated, where R is given by the relation

$$R = \frac{(\gamma + 1)^2 f^2 \times 10^7 W}{2\rho c^3 f_0^2 A}, \tag{1}$$

in which f is the frequency of the fundamental, f_0 the cut-off frequency of the horn, c the velocity of sound, ρ the density of air, and A the area of the throat of the horn, all expressed in c.g.s. units. It may be noted that the intensity of the harmonic increases with the ratio of the frequency to the cut-off frequency of the horn; this is another argument against attempting to cover too wide a range of frequencies with a single horn. In Fig. 1 it is shown that in the region of 200 c.p.s. the orchestra gives peak powers of about 10 watts. If, therefore, 30 watts be set as the limit of power that the horn is to deliver at 200 c.p.s., 32 c.p.s. as the cut-off frequency of the horn, and 30 db below the fundamental be assumed as the limit of tolerance of a second harmonic, from equation (1) a throat diameter of about 8 inches is determined.*

If the radiation resistance at the throat of a horn is not to vary appreciably with frequency, the mouth opening must be a substantial fraction of a wave-length. This condition calls for an unusually large horn if frequencies down to 40 c.p.s. and below are to be transmitted. However, the effect of variations in radiation resistance on sound output can be kept down to a relatively small value if the receiving unit is properly designed. This will be explained in the

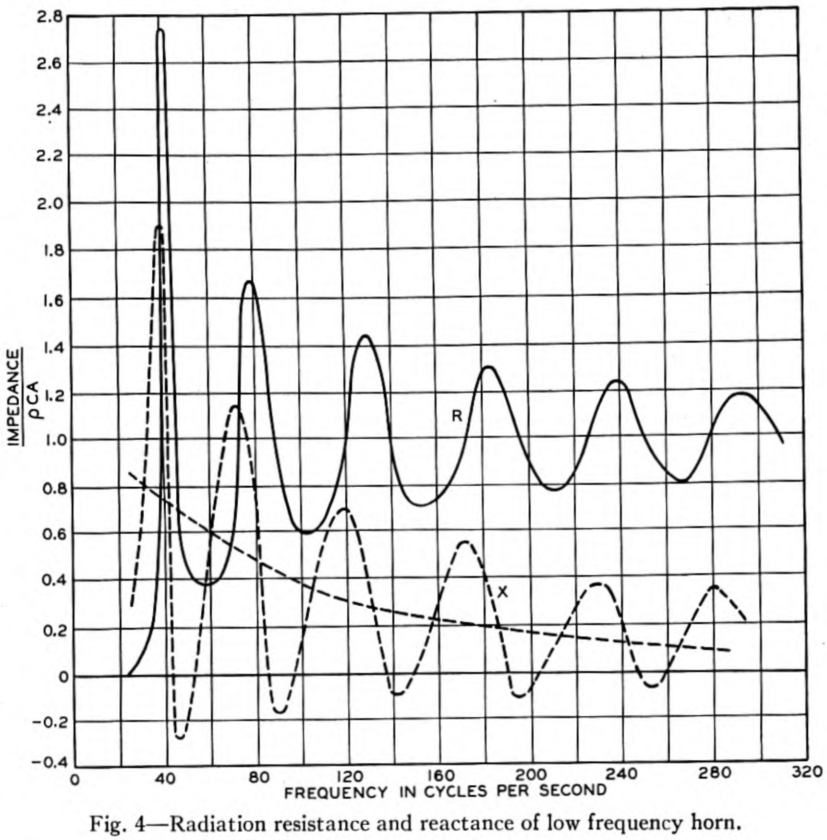

Fig. 4—Radiation resistance and reactance of low frequency horn.

next section. The low frequency horn used in these reproductions has a mouth opening of about 25 square feet. As computed from well-known formulas [7] for the exponential horn the impedance of this horn

* Since the original publication of this paper, experimental data have been obtained which indicate a second harmonic generation in horns 6 or more db below the value shown by Rocard's equation.[9]

with a throat diameter of 8 inches is shown in Fig. 4. These curves were computed under the assumption that the mouth of the horn is surrounded by a plane baffle of infinite extent, a condition closely approximated if the horn rests on a stage floor.

Low Frequency Receiving Unit

When a moving coil receiving unit, coupled to a horn, is connected to an amplifier having an output resistance equal to $n - 1$ times the damped resistance R of the driving coil, it can easily be shown that the sound power output is

$$P = \frac{\left(\dfrac{EBLT}{nR}\right)^2 r \times 10^{-9}}{\left[T^2 r + \dfrac{B^2L^2 \times 10^{-9}}{nR}\right]^2 + [x_d + T^2 x]^2} \text{ watts,} \tag{2}$$

where E is the open circuit voltage of the amplifier, L the length of wire in the receiver coil, T the ratio of the area of the diaphragm to the throat area of the horn, $r + jx$ the throat impedance of the horn, and x_d the mechanical reactance of the diaphragm and coil, the mechanical resistance of which is assumed to be negligibly small. From Fig. 4 it may be seen that the mean value of x increases as the frequency decreases to a value below 40 c.p.s., and that x is smaller than r except at the very lowest frequencies. If, therefore, the stiffness of the diaphragm be adjusted so that x_d is equal to T^2 times the mean value of x at 40 c.p.s., the second term in the denominator may be neglected without much error because it will have but little effect upon the sound output except at the higher frequencies, where the mass reactance of the coil and diaphragm may have to be taken into account.

If minimum variations in sound output are desired for variations in r,

$$\frac{B^2L^2 10^{-9}}{nRT^2} = r_0, \tag{3}$$

where r_0 is equal to the geometric mean value of r, which is approximately equal to $A\rho c$.

If α is the ratio of the resistance at any frequency to the mean value, and if the second term in the denominator is neglected, equation (2) becomes

$$P = \frac{E^2}{nR}\frac{\alpha}{(1 + \alpha)^2}\cdot \tag{4}$$

In Fig. 4 it is shown that above 35 c.p.s. α has extreme values of 2.75 and 0.36, at which points there will be minimum values in P, but these minimum values will not lie more than 1 db below the maximum values. Hence, if the receiver satisfies the condition of equation (3), the extreme variations in the sound output will not exceed 1 db, although the horn resistance varies by a factor of 7.5. Also it may be stated here that when the condition of equation (3) is satisfied the horn is terminated at the throat end by a resistance equal to the surge resistance of the horn. Thus equation (3) establishes a condition of minimum values in the transient oscillations of the horn.

The mean motional impedance of the loud speaker is $\frac{B^2L^2 \times 10^{-9}}{T^2 r_0}$, which, from equation (3), is equal to nR. The condition of equation (3) therefore specifies that the efficiency of the loud speaker shall be $\frac{n}{n+1}$. The maximum power that an amplifier can deliver without introducing harmonics exceeding a specified value is a function of the impedance into which it operates. Therefore, to obtain the maximum acoustic power for a specified harmonic content, the load impedance should have the value for which the product of the loud speaker efficiency and the power capacity of the amplifier has a maximum value. This optimum value of load impedance for the amplifier and loud speaker used in the Philadelphia-Washington experiments was found to be about 2.25 times the output impedance of the amplifier; the corresponding value of n then is 2.6 and the required efficiency 72 per cent. For best operating condition a definite value of receiver efficiency thus is specified.

The receiver may be made to satisfy the foregoing conditions regardless of the value of T, the ratio of diaphragm area to throat area. The area of the diaphragm has, however, a definite relation to the maximum power that the receiver can deliver at the low frequencies. The peak power delivered by the receiver is equal to $T^2 \alpha r_0 \xi^2 \omega^2 \times 10^{-7}$ peak watts where ξ is the maximum amplitude of motion of the diaphragm. Figure 1 shows that in the region lying between 40 and 60 c.p.s., peak powers reach a value of from 1 to 2 watts. However, the low frequency tones of an orchestra are undesirably weak and may advantageously be reproduced at a relatively higher level. Therefore it was decided to construct the loud speaker to be able to deliver 25 watts in this region.

As the coil moves out of its normal position in the air gap, the force factor varies. Harmonics thus will be generated, the intensities of which increase with increasing amplitude. A limit to the maximum value of the amplitude ξ thus is set by the harmonic distortion that one is willing to tolerate. In this receiver the maximum value of ξ

was taken equal to 0.060 in. Figure 4 shows that $\alpha\omega^2$ has a minimum value at about 50 cycles, where α is equal to about 0.4. These values give a ratio of 4.5 for T.

Inasmuch as $R = \frac{\sigma L^2}{v}$, where σ is the resistivity of the wire used for the coil and v the volume of the coil, from equation (3) is obtained

$$B^2 v = n\sigma T^2 r_0 10^9. \tag{5}$$

The first member gives the total magnetic energy that must be set up in the region occupied by the driving coil. This value is fixed by the fact that all factors in the second member are specified. The same performance is obtained with a small coil and high flux density as with a large coil and low flux density, provided B^2v is held fixed, but the coil in any case should not be made so small that it will be incapable of radiating the heat generated within it without danger of overheating, nor so large that the mass reactance of the coil will reduce the efficiency at the higher frequencies.

This receiver unit, when constructed according to the above principles and when connected to an amplifier and a horn in the specified manner, should be capable of delivering power 3 or 4 times that delivered by the orchestra in the frequency region lying between 35 and 400 c.p.s., with an efficiency of about 70 per cent, and with a variation in sound output for a given input power to the amplifier of not more than 1 db throughout this range.

The High Frequency Horn

It is well known that a tapered horn of the ordinary type has a directivity which varies with frequency. Sound of low frequency is projected through a relatively large angle. As the frequency is increased this angle decreases progressively until, at frequencies for which the wave-length is small compared with the diameter of the mouth opening, the sound beam is confined to a very narrow angle about the axis of the horn.

If we had a spherical source of sound (i.e., a source consisting of a sphere, the surface of which has a radial vibratory motion equal in phase and amplitude at every point of the surface), sound would be radiated uniformly outward in all directions; or, if we had only a portion of a spherical surface over which the motion is radial and uniform, uniform sound radiation still would prevail throughout the solid angle subtended at the center of curvature by this portion of the sphere, provided its dimensions were large compared with the wave-length. Throughout this region the sound would appear to originate at the center of curvature. Hence, for the ideal distribution of a spherical source within a region to be defined by a certain solid angle, it is necessary and sufficient that the radial motion be the same in amplitude and phase over the part of a spherical surface intercepted by the angle and having its center of curvature at the vertex and located at a sufficient distance from the vertex to make its dimensions large compared with the wave-length. If, further, these conditions are satisfied for this surface at all frequencies, all points lying within the solid angle will receive sound of the same wave form. A horn was designed to meet these requirements for the high frequency band.

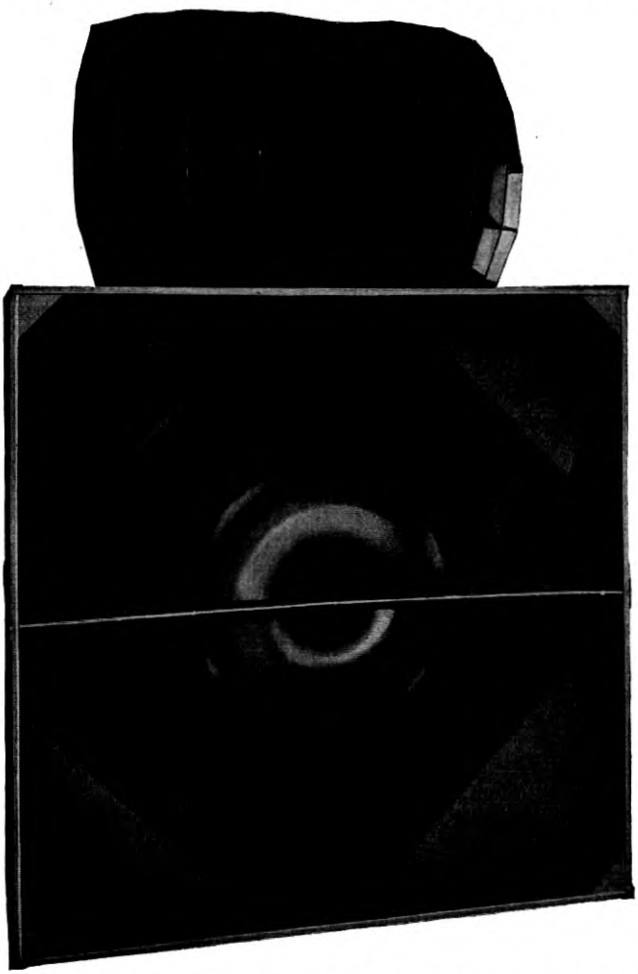

Fig. 5—Special loud speaker developed for auditory perspective experiment.

The horn, shown in the upper part of Fig. 5, comprises several separate channels, each of which has substantially an exponential taper. Toward the narrow ends these channels are brought together with their axes parallel, and are terminated into a single tapered tube which at its other end connects to the receiver unit. Sound from the latter is transmitted along the single tube as a plane wave and is divided equally among the several channels. If the channels have the same taper, the speed of propagation of sound in them is the same. The large ends are so proportioned and placed that the particle motion of the air will be in phase and equal over the mouth of the horn. This design gives a true spherical wave front at the mouth of the horn at all frequencies for which the transverse dimensions of the mouth opening are a large fraction of a wave-length.

As the frequency is increased, the ratio of wave-length to transverse width of the channels becomes less, and the sound will be confined more and more to the immediate neighborhood of the axis of each channel. The sound then will not be distributed uniformly over the mouth opening of the horn, but each channel will act as an independent horn. To have a true shperical wave front up to the highest frequencies, the horn would have to be divided into a sufficient number of channels to make the transverse dimension of each channel small compared with the wave-length up to the highest frequencies. If it is desired to transmit up to 15,000 c.p.s., it is not very practical to subdivide the horn to that extent. Both the cost of construction and the losses in the horn would be high if designed to transmit also frequencies as low as 200 c.p.s., as is the case under consideration. However, it is not important that at very high frequencies a spherical wave front be established over the whole mouth of the horn. For this frequency region it is perfectly satisfactory to have each channel act as an independent horn, provided that the construction of the horn is such that the direction of the sound waves coming from the channels is normal to the spherical wave front.

The angle through which sound is projected by this horn is about 60 degrees, both in the vertical and in the horizontal direction. For reproducing the orchestra two of these horns, each with a receiving unit, were used. They were arranged so that a horizontal angle of 120 degrees and a vertical angle of 60 degrees were covered. These angular extensions were sufficient to cover most of the seats in the hall with the loud speaker on the stage. The vertical angle determines to a large extent the ratio of the direct to the indirect sound transmitted to the audience. The vertical angle of 60 degrees was chosen purely on the basis of judgment as to what this ratio should be for the most pleasing results.

The High Frequency Receiving Unit

In the design of the low frequency receiver one of the main objectives was to reduce to a minimum the variations in sound transmission resulting from variations in the throat impedance of the horn. However, the high frequency horn readily can be made of a size such that the throat resistance has relatively small variations within the transmitting region. On the other hand, whereas the diameter of the diaphragm of the low frequency unit is only a small fraction of the wave-length, that of the high frequency unit must be several wavelengths at the higher frequencies in order to be capable of generating the desired amount of sound. Unless special provisions are made there will be a loss in efficiency because of differences in phase of the sound passing to the horn from various parts of the diaphragm. The high frequency receiver therefore was constructed so that the sound generated by the diaphragm passes through several annular channels. There are enough of these channels to make the distance from any part of the diaphragm to the nearest channel a small fraction of a wave-length. These channels are so proportional that the sound waves coming through them have an amplitude and phase relation such that a substantially plane wave is formed at the throat of the horn.

In the appendix it is shown that, for the higher frequencies where the impedance of the horn may be taken as equal to ρc times the throat area and for the type of structure adopted, the radiation resistance is equal to

$$\rho c a T^2 \left[\frac{1}{k^2 h^2 T^2 + k^2 l^2 \cot^2 kl} \right] \tag{6}$$

and the reactance

$$-j\frac{\rho c a}{kh} T \left[1 - \frac{1}{kl \cot kl + \left(\frac{hT}{l} \right)^2 kl \tan kl} \right]. \tag{7}$$

where a is the area of the throat of the horn, T the ratio of the area of the diaphragm to the throat area, $k = \omega/c$, and the other designations are those indicated in Fig. 11. At the lower frequencies the resistance is $T^2 r$ and the reactance $T^2 x$, where r and x are, respectively, the resistance and reactance of the throat of the horn.

Equation 6 shows that at a given frequency, other conditions remaining the same, the radiation resistance will have a maximum value when l is approximately equal to $\pi/2k = c/4f$. In Fig. 6 the resistances as computed from equation (6) are plotted as a function of frequency for several values of h/w. It is seen from these curves that the resistance at the higher frequencies is determined very largely by the relation of h/w but is independent of it at the lower frequencies, where it is equal to $\rho ca T^2$. At the lower frequencies where the mechanical impedance of the diaphragm is negligible, the efficiency, as was the case for the low frequency receiver, depends

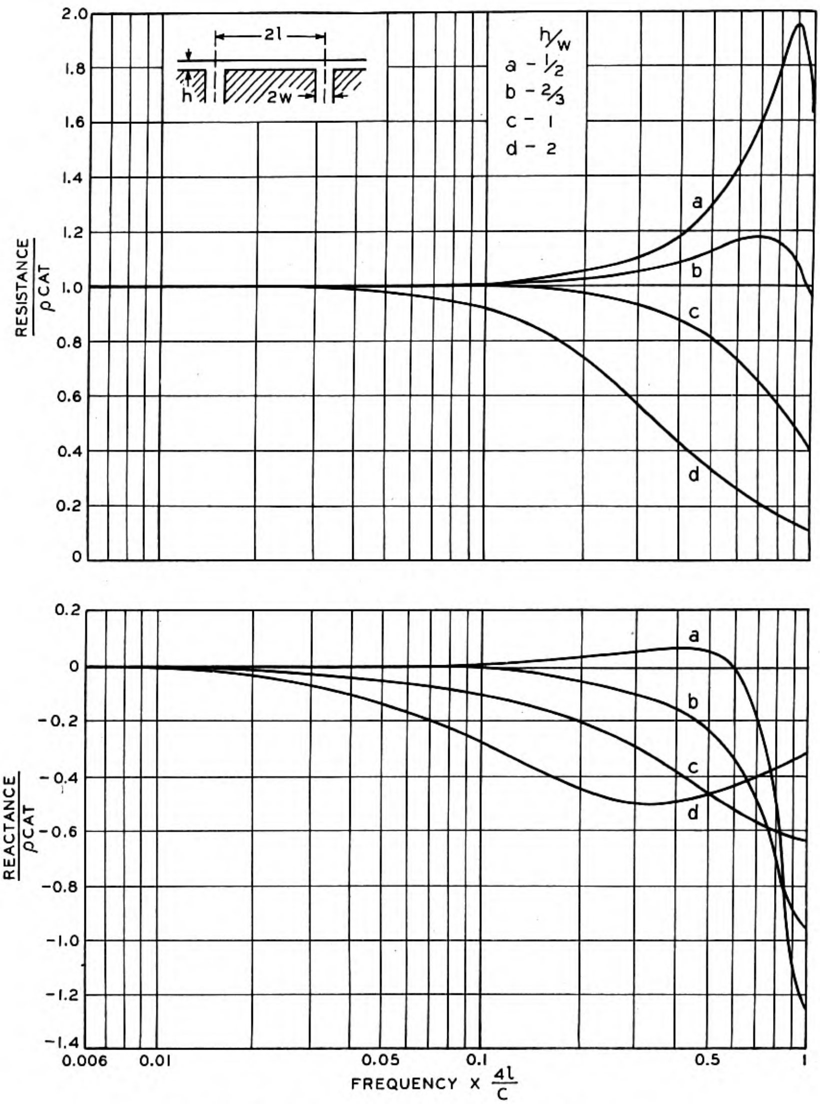

Fig. 6—Load impedance of speaker diaphragm.

upon the value of B^2v where v is the volume of the coil, but at the higher frequencies the efficiency decreases with increasing mass of the coil. It is advantageous, therefore, to keep v small and to make B as large as is practically possible. Values were selected to give the receiver an efficiency of 55 percent at the lower frequencies. For these conditions the relative sound power output was computed by equation (2) on the assumption that the receiver was connected to an amplifier having an output impedance equal to 0.45 times that of the receiver

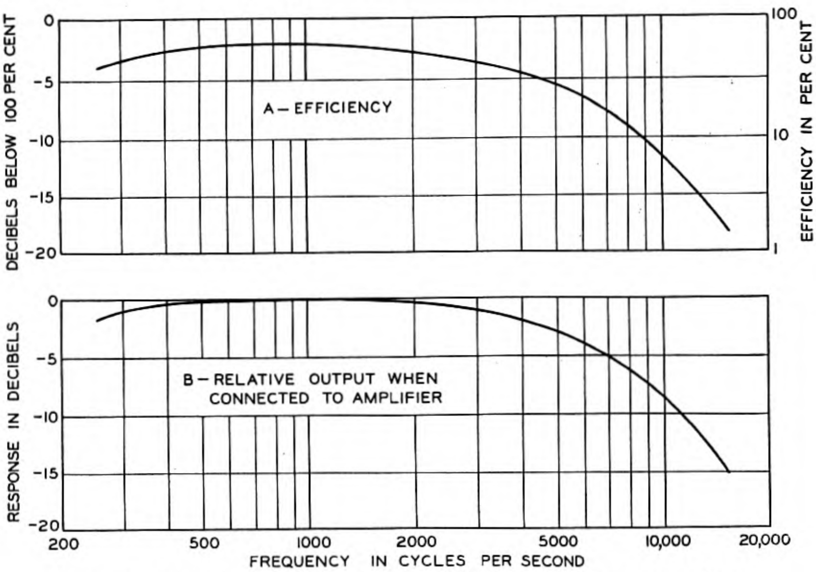

Fig. 7—Relative computed sound output of high frequency receiver.

at the lower frequencies. Figure 7 shows the values so obtained. Corresponding values obtained experimentally when the receiver was connected to the horn previously described are shown in Figs. 8 and 9, where the sizes of the rooms in which the values were obtained were, respectively, 5000 and 100,000 cubic feet. Both of these curves differ considerably from the computed curve, particularly as regards loss at high frequencies. The curve of Fig. 8 shows less, and that of Fig. 9 more, loss at high frequencies. The computed curve, however, refers to the total sound output, whereas the measured curves give average values of sound intensity in a certain part of the room, values dependent upon the acoustic characteristics of the room.

The number of high frequency receivers that must be used for each transmitting channel is governed largely by the amount of power that the system is to deliver before harmonics of an objectionable intensity

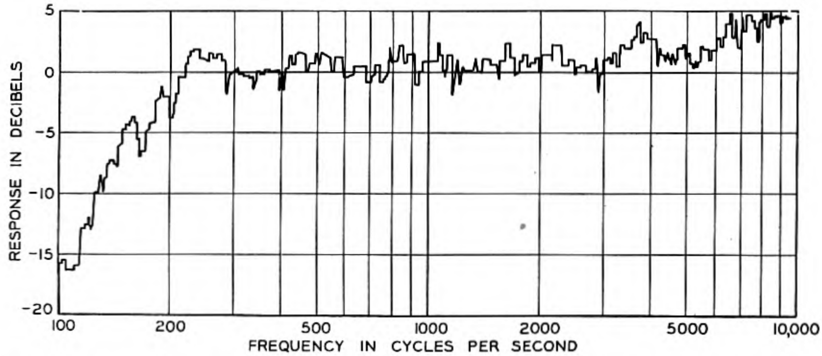

Fig. 8—Output-frequency characteristic of high frequency receiver as measured in a small room.

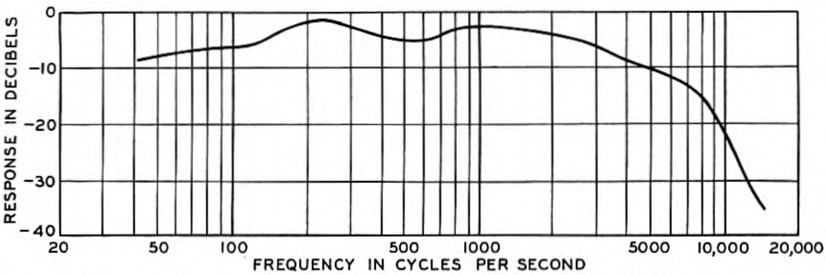

Fig. 9—Output-frequency characteristic of combined low and high frequency receivers as measured in a large room.

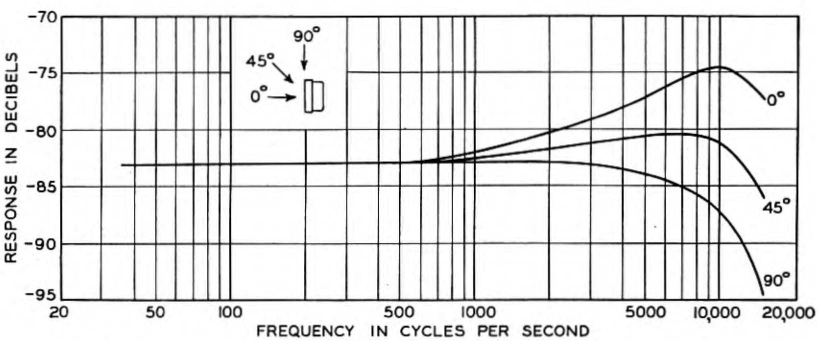

Fig. 10—Output-frequency characteristic of moving coil microphone.

are introduced. The generation of harmonics in a horn when transmitting waves of large amplitude already has been discussed. Let it suffice here to say that, for a given percentage harmonic distortion, the power that can be transmitted through the horn is proportional to the area of the throat and inversely proportional to the square of the ratio of the frequency to the cut-off frequency.

Inasmuch as the moving coil microphones used for the transmission of music in acoustic perspective have been described previously [8] they will not be discussed here at length. Their frequency response characteristic as measured in an open sound field for several different angles of incidence of the sound wave on the diaphragm are shown in Fig. 10 where it is seen that the response at the higher frequencies becomes less as the angle of incidence is increased. In general, this is not a desirable property, but with the instruments as used in this experiment the sound observed as coming from each loud speaker is mainly that which is picked up directly in front of each microphone; sound waves incident at a large angle do not contribute much.

At certain times the sound delivered by the orchestra is of very low intensity. Therefore it is important that the microphones have a sensitivity as great as possible, so that the resistance and amplifier noises may readily be kept down to a relatively low value. At 1,000 c.p.s. these microphones, without an amplifier, will deliver to a transmission line 0.05 microwatt when actuated by a sound wave having an intensity of 1 microwatt per square centimeter. This sensitivity is believed to be greater than that of microphones of other types having comparable frequency response characteristics, with the possible exception of the carbon microphone.

APPENDIX

Load Impedance of a Diaphragm Near a Parallel Wall with Slot Openings

First assume a diaphragm and a parallel wall of infinite extent separated by a distance h, and that the wall is slotted by a series of equally spaced openings as shown in Fig. 11. From symmetry it is known that when the diaphragm vibrates there will be no flow perpendicular to the plane of the paper or across the planes indicated by the dotted lines. Therefore only one portion of unit width, such as *abcdef* need be considered. Let the x and y reference axes be located as shown. If the general field equation

$$\frac{\partial^2 \varphi}{o x^2} + \frac{\partial^2 \varphi}{\partial y^2} + k^2 \varphi = 0 \qquad (8)$$

is applied when the diaphragm has a normal velocity equal $\dot{\xi}e^{i\omega t}$ the following boundary conditions are obtained:

When
$$x = 0, \quad \partial\varphi/\partial x = -\dot{\xi},$$
$$x = h, \quad \partial\varphi/\partial x = 0,$$
$$y = 0, \quad \partial\varphi/\partial y = 0,$$

and when $y = l$, the pressure is equal to the product of acoustic impedance and volume velocity or

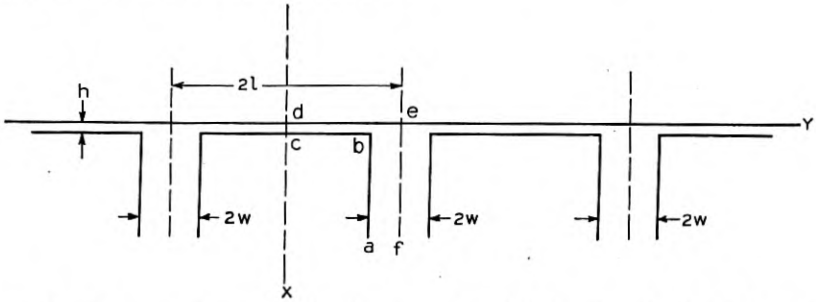

Fig. 11—Schematic diagram of diaphragm and parallel slotted wall of infinite length.

$$\frac{\rho}{h}\int_0^h \left(\frac{\partial\varphi}{dt}\right)_{y=l} dx = \frac{c\rho}{w}\int_0^h \left(-\frac{\partial\varphi}{\partial y}\right)_{y=l} dx$$

where φ is the velocity potential, $k = \omega/c$, and c is the velocity of sound.

The appropriate solution of equation (8) then is

$$\varphi = \frac{\dot{\xi}}{k}\left[\frac{\cos ky}{kh\left(\cos kl + i\dfrac{h}{w}\sin kl\right)} - \frac{\cos k(x-h)}{\sin kh}\right].$$

The average reacting force per unit area of the diaphragm is

$$\frac{ik\rho c}{l}\int_0^l (\varphi)_{x=0}\, dy$$

Thus, for the impedance per unit area, which is equal to the force divided by the velocity, is obtained

$$\frac{\rho c l}{w}\left\{\left[\frac{\sin^2 kl}{k^2l^2}\,\frac{1}{\cos^2 kl + \left(\dfrac{h}{w}\right)^2 \sin^2 kl}\right] - j\frac{w}{h}\left[\frac{\dfrac{kh\cos kh}{\sin kh}kl - \dfrac{\sin kl\cos kl}{\cos^2 kl + \left(\dfrac{h}{w}\right)^2\sin^2 kl}}{k^2l^2}\right]\right\} \equiv r' + jx'.$$

In all practical types of loud speakers $kh \cos kh/\sin kh$ would be very nearly equal to 1; then

$$r' = \frac{\rho cl}{w}\left[\frac{1}{k^2l^2\left(\left(\frac{h}{w}\right)^2 + \cot^2 kl\right)},\right]$$

$$x' = -\frac{\rho cl}{h}\left[\frac{kl - \dfrac{1}{\cot kl + \left(\dfrac{h}{w}\right)^2 \tan kl}}{k^2l^2}\right].$$

If the total area of the diaphragm is A and that of the corresponding channels a, then $A/a = l/w$, approximately, and the total impedance becomes

$$r = \frac{\rho c A^2}{a}\cdot\frac{1}{\left(\frac{kh}{a}\right)^2 A^2 + k^2l^2\cot^2 kl},$$

$$x = -j\frac{\rho cA}{kh}\left[1 - \frac{1}{kl\cot kl + \left(\frac{h}{l}\frac{A}{a}\right)^2 kl\tan kl}\right].$$

REFERENCES

1. "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestra," L. J. Sivian, H. K. Dunn, and S. D. White. *Jour. Acous. Soc. Am.*, v. 2, Jan. 1931, p. 330.
2. "Audible Frequency Ranges of Music, Speech, and Noise," W. B. Snow. *Jour. Acous. Soc. Am.*, v. 3, July 1931, p. 155.
3. Backhaus, *Zeits. f. Tech. Physik*, v. 9, 491, 1928.
4. "Engineering Acoustics," L. E. C. Hughes, p. 47. Benn, London.
5. W. J. Albersheim and J. P. Maxfield, similar relations were presented in a paper before the Acoustical Society in May 1932.
6. "Sur la Propagation des Ondes Sonores d'Amplitude Finie," M. Y. Rocard. *Comptes Rendus*, Jan. 16, 1933, p. 161.
7. "Theory of Vibrating Systems and Sound," Crandall. P. 163 ff. D. Van Nostrand, New York.
8. "Moving Coil Telephone Receivers and Microphones," E. C. Wente and A. L. Thuras. *Jour. Acous. Soc. Am.*, v. 3, 1931, p. 44.
9. "Extraneous Frequencies Generated in Air Carrying Intense Sound Waves," A. L. Thuras, R. T. Jenkins and H. T. O'Neil. To be presented at mtg. of Acous. Soc. Am., April 30–May 1, 1934.

Amplifiers*

By E. O. SCRIVEN

Appreciable care is required in the design of a system which must amplify with great fidelity practically the whole range of audible frequencies and be capable of delivering a high level while at the same time providing a wide volume range. Some of the problems involved are discussed, particularly as applying to the equipment used in the reproduction in Washington, D. C., of the Philadelphia Symphony Orchestra playing in Philadelphia.

VACUUM tube amplifiers have been closely identified with the extension of the channels of communication since, with completion of the initial transcontinental telephone line 20 years ago, they first enabled New York to converse with San Francisco. There are now thousands of audio frequency amplifiers in telephone circuits and in sound picture theaters, public address systems, and other similar services as well as in the millions of radio receiving sets.

Along with the extension of the field of usefulness of audio amplifiers there has been continuing progress toward more faithful reproduction, better transmitters, better receivers, and better amplifiers. Those first telephone repeaters, although quite adequate for their immediate purpose, transmitted a frequency band only a few octaves wide. Very few radio sets even now cover a range above 3,000 c.p.s. without distortion, and the most up-to-date sound picture installation rarely can be depended upon for accurate reproduction of frequencies above 7000 or 8000 c.p.s. The requirements as to frequency range and freedom from distortion for any particular service are, in the last analysis, determined by public demand.

However, when one undertakes to reproduce an orchestra like the Philadelphia Symphony and to reproduce it in such a manner as to satisfy the critical ear of the director, or that of the devotee of symphonic concerts, one has to provide something out of the ordinary in audio amplifiers.

In his paper, which forms a part of this symposium, Dr. Fletcher has pointed out that only the elimination of those frequencies below 40 c.p.s. and those above 15,000 c.p.s. produces no detectable difference in the reproduction of symphonic music. This, then, is the

* Fourth paper in the Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering*, January, 1934.

frequency spectrum that the amplifier must be designed to handle. Also, it is important that there shall be uniform amplification of all parts of the frequency range and that no extraneous frequencies shall be introduced.

Of importance commensurate with the distortionless amplification of the complete frequency range of the orchestra is the provision of an equivalent volume of sound. The amplifier must be capable of supplying to the loud speakers without distortion an amount of energy that will produce a sound volume at least equivalent to that produced by the orchestra (the Philadelphia-Washington installation was designed to produce about 10 times this amount). And equally important, the amplifier must be so free from internal disturbances and from self-induced electrical fluctuations that the softest music, the weakest input to the microphone, can be reproduced without appreciable background noise. According to Fletcher the ratio of the heaviest playing of a large orchestra such as the Philadelphia Symphony Orchestra to the softest music such as that of a violin is about 10,000,000 to 1, or 70 db. Thus it is required that any noise be at least 75 db below the loudest tones; that is, there must be at least a 75-db volume range.

The sources of noise may be divided into 2 groups. In the first group are included the 60-cycle alternating current power supply, vibration or jar of mechanically unstable vacuum tubes, contact and thermoelectric potentials, and similar disturbances, which may be reduced to practically any degree depending upon the lengths to which one is willing to go to reduce them. In the second group are those electronic irregularities intimately associated with the operation of the vacuum tube and which depend somewhat upon the design, manufacture, and method of operation of the vacuum tube; and which, when sufficiently amplified and fed into a loud speaker, may be heard as noise. In general, the maximum volume range of an amplifier is reached when all other disturbances are reduced to the level of this tube noise.

It is evident, then, that under ordinary circumstances the limiting volume range of an amplifier is a function of the amount of amplification following the first tube. In other words, the magnitude of the signal voltage with respect to the noise voltage in the plate circuit of the first tube in a multistage amplifier determines the limiting volume range obtainable with that amplifier.

It will appear that in a sound reproduction system a highly efficient microphone simplifies the amplifier volume range requirements, and that loud speakers of high efficiency reduce the volume required from the amplifier.

Perhaps it is in order to inquire as to what makes an amplifier free from frequency distortion over a wide range. The answer might well be: attention to impedance relations. A compact, efficient amplifier requires several pieces of reactive apparatus such as transformers, retardation coils, and capacitors. One must remember that an inductance of one henry is equivalent to an impedance of 250 ohms at 40 c.p.s. but that it is nearly 100,000 ohms at 15,000 c.p.s.; that the grid circuit of the vacuum tube is not actually an open circuit even though the grid is maintained negative with respect to the cathode, but has a reactance which becomes important at high frequencies or with large ratio input transformers. Many years of development in this field have advanced the art to the point where transformers transmitting extremely wide bands now can be designed. The commercial production of such designs requires rigid inspection including shop transmission measurements under the actual conditions of use. The transformer must be designed for the particular type of vacuum tube with which it is to be used. First, however, the tube must be designed to permit its use under the proposed conditions and then it must be manufactured to close limits, every tube of a type like every other tube of that type.

This is, then, the general requirement for a wide frequency range amplifier: (1) attention to impedance relations; (2) meticulous design of each component for the particular job it has to do, and rigid inspection to insure that it does that job.

One might suppose that when the tube designer and the coil designer each had done his part the job was done. Such is not the case. The various pieces of apparatus have to be gathered together into a unit (often a current supply set for supplying anode, cathode, and grid potential is assembled with the amplifier) and out of this electrical and physical association is apt to arise "feed-back" and "noise."

When there is coupling between two parts of the amplification circuit which are at different potential or different phase there is feed-back. Feed-back sometimes is employed designedly to modify an amplifier characteristic, but, feed-back which may arise as a result of a particular arrangement of apparatus or wiring ordinarily will cause more or less severe frequency distortion. It may be induced due to stray electromagnetic or electrostatic fields, which must be eliminated by rearrangement of apparatus or by shielding; or it may be caused by common circuit impedance, requiring circuit modifications. In general, a low gain amplifier or one with limited frequency range presents no feed-back problems, but a study of a high-gain wide-range equipment usually is necessary in order to determine the best arrangement. Often modifications of tentative circuit or apparatus must be made to obtain satisfactory operation.

The provision of a volume range of some 75 db on an energy basis became largely a matter of the suppression of a.-c. hum. The low inherent electronic noise effect of the Western Electric No. 262A vacuum tube and the relatively high level from the microphones kept electronic tube noise well in the background. Careful and in some cases rather elaborate shielding of audio transformers and leads and the segregation of the 60-cycle power equipment coupled with the use of vacuum tubes having indirectly heated cathodes and specially designed to have small stray fields prevented a.-c. hum trouble in the early stages. However, the Western Electric No. 242A vacuum tubes used in the push-pull final stage have filamentary cathodes, and when such tubes have raw a.-c. filament supply, a very appreciable 120-cycle component appears in the space current. Although theoretically in a perfectly balanced push-pull amplifier this component would be eliminated, in practice an exact balance cannot be obtained. As a final step in noise elimination, advantage was taken of the fact that each channel employed two amplifiers in parallel. Under such conditions and with proper phasing of the power supply to the two amplifiers the net a.-c. noise output of the two amplifiers in parallel will be less than that of either one alone.

Having reduced feed-back and noise to tolerable values, it remains to determine the operating conditions for maximuum output. The vacuum tube is not strictly a linear device, but, when properly used, the total harmonic content can be held to a low figure. For a high quality system the total harmonics produced in the system should not exceed one per cent of the fundamental. This requires that impedance and potential relations in the vacuum tubes should be adjusted to give approximately linear operation; and also that the design of audio transformers, particularly those carrying considerable levels, must be scrutinized carefully to insure that they operate over an essentially linear portion of the magnetization curve of the core material.

An instrument really essential to the design of high quality amplifiers is a high sensitivity harmonic analyzer that is capable of quickly and accurately resolving a complex wave into its simple components. By this means the effect of variations in circuit relations can be evaluated and the optimum condition for maximum distortionless power output determined.

It may be desirable at this point to examine the make-up of the audio amplification system used in the Philadelphia-Washington experiments. It should be noted that the arrangement of equipment

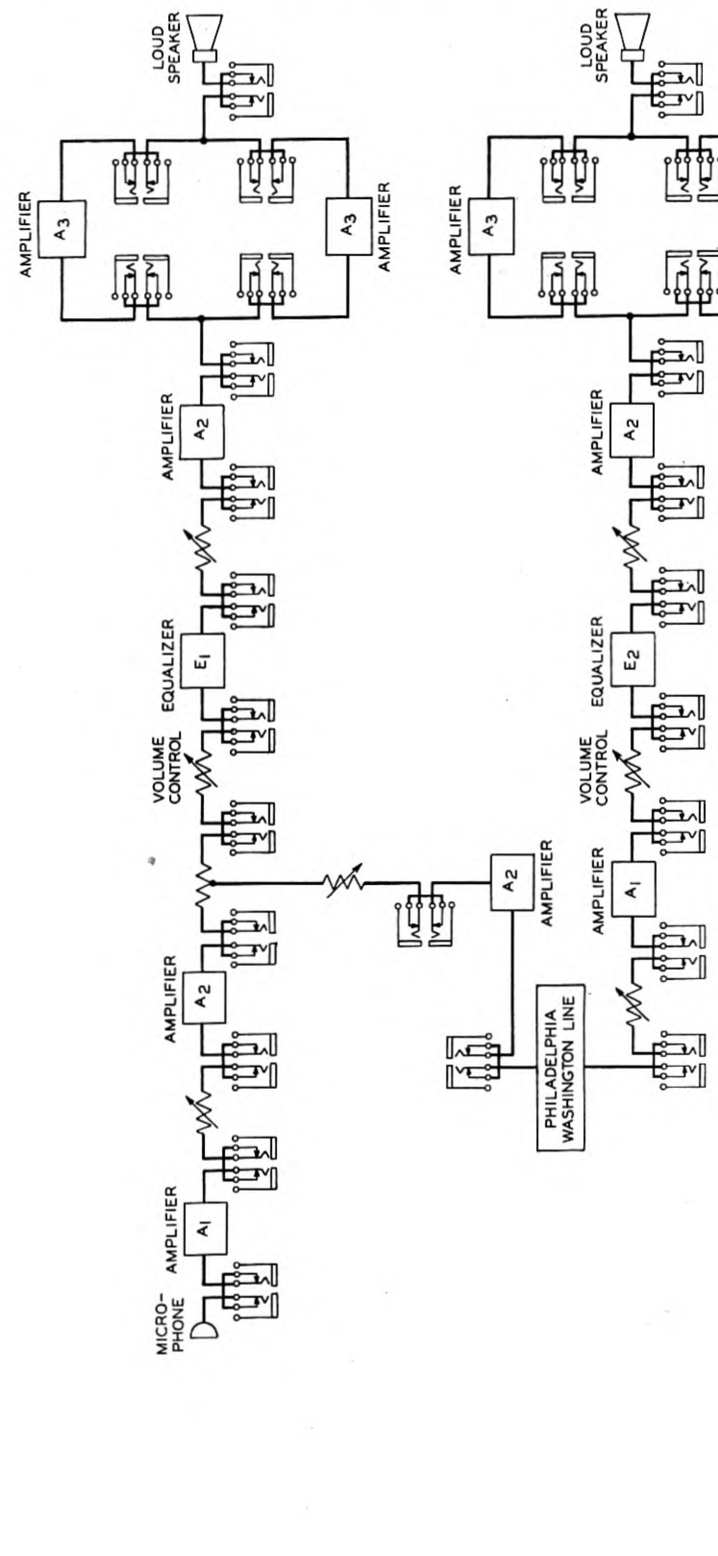

Fig. 1—Schematic diagram of the amplification system used in conveying Philadelphia symphonic music to Constitution Hall in Washington, D. C., and there reproducing it in auditory perspective.

provided for simultaneous reproduction at both Philadelphia and Washington. There were three complete and essentially equivalent channels of equipment actually in use and a fourth complete channel held in reserve as a spare.

Several stages of so-called voltage amplification were required preliminary to the final or power stage. There is, of course, no essential difference between a voltage amplifier and a power amplifier, the term "voltage amplifier" being applied to those preliminary stages of an amplification system the function of which is so to amplify the output of the pick-up device as to supply adequate driving voltage to the grids of the power stage. Theoretically, inasmuch as no energy is absorbed in the ideal grid circuit, this voltage increase might be supplied entirely by a high ratio input transformer. However, there are practical difficulties to the design of such a single stage amplifier and therefore multistage vacuum tube amplification is employed.

As a matter of convenience the voltage amplification for this system was obtained through the use of several separate amplifier units in tandem. This arrangement not only enabled the ready replacement of any unit of the system in case of failure, but it also facilitated the insertion of a pad, control potentiometer, or other network at any desired point. Several of these devices were required, and of course each introduced a loss. Thus the gross amplification of the system used for reproduction at Philadelphia was approximately 160 db and for Washington 240 db, although the actual difference in level between microphone output and loud speaker input was but from 80 to 90 db.

The general scheme of the amplification system is shown in Fig. 1. A_1 is a single-stage, single-tube Western Electric No. 80A amplifier slightly modified to meet the particular conditions of use; it has a gain of 30 db, and employs a Western Electric No. 262A vacuum tube. This tube has an equipotential cathode, the heater being operated on 10-volt 60-cycle alternating current and the anode being supplied from rectified alternating current. A_2 is a 2-stage amplifier having a single Western Electric No. 262A vacuum tube in the first stage and push-pull Western Electric No. 272A tubes in the second stage. It has a gain of 50 db. The cathodes of the tubes are energized with low-voltage 60-cycle alternating current and the anodes with rectified alternating current. A_3, the final or power amplifier, is a single stage amplifier employing two Western Electric No. 242A vacuum tubes in parallel on each side of a push-pull circuit, thus having four tubes per amplifier. Two of the A_3 amplifiers were used in parallel on each channel, and were capable of supplying 60 watts each, or a total of 120 watts, to the loud speakers. These are r.m.s. values. The instantaneous peaks of power of course could equal twice this value, or 720 watts, for the three channels. E_1 and E_2 are equalizers to compensate for any amplitude distortion that would cause a listener to obtain a different tone effect from the loud speakers than he would from the

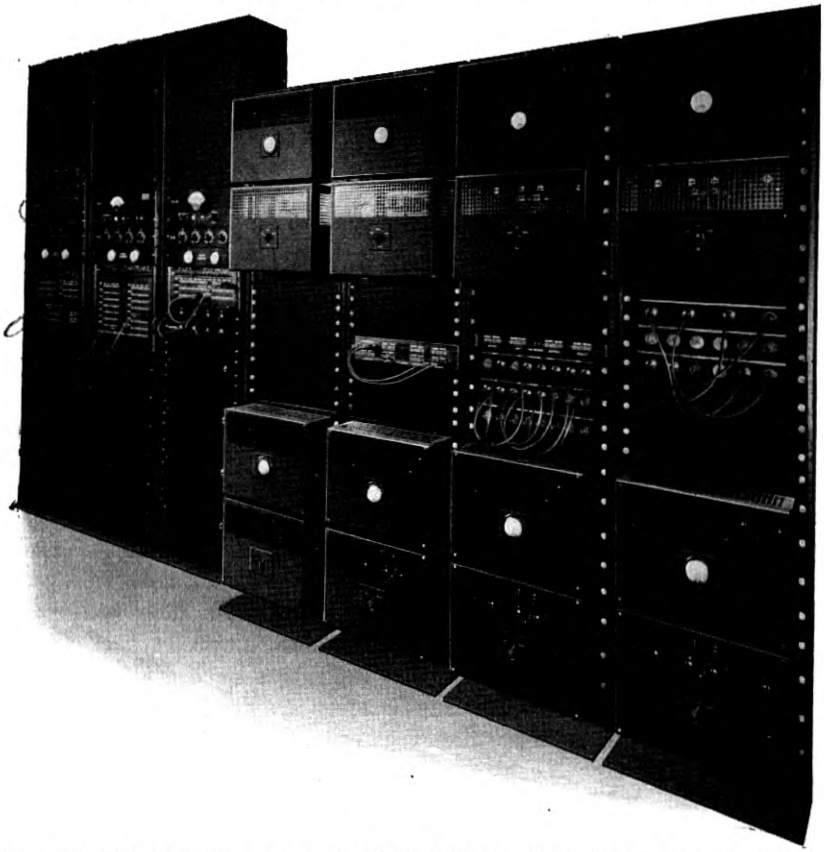

Fig. 2—Amplifying equipment used at Philadelphia. The taller racks are 8 ft. high and contain A_1 and A_2 amplifiers, volume indicators, and various controls.

actual orchestra. These equalizers are loss networks and principally equalize for the acoustic characteristic of the loud speakers in the particular hall, but they are placed in a low energy part of the amplification circuit so as not to waste the energy of the final power stage.

In view of the inclusion of the equalizers in the amplification system, and particularly because of the fact that the amplification of the A_3 amplifier deliberately was made to increase with frequency in order to compensate in part for acoustic losses in the overall system, the actual amplification-frequency curves of the amplifiers are of little importance. The equalizers of the system are discussed in the paper by Bedell and Kerney.

Transmission Lines*

By H. A. AFFEL, R. W. CHESNUT and R. H. MILLS

Describing methods whereby high quality sound reproduction in auditory perspective can be accomplished over long distances, this discussion centers largely upon a description of the exact technique employed in providing communication transmission circuits for the Philadelphia-Washington demonstration. Problems that might be involved in carrying out such transmission on a more widespread scale also are touched upon.

MICROPHONES have been described that will pick up without noticeable distortion all the sounds given forth by a symphony orchestra. Loud speakers and amplifiers also have been described that will accurately reproduce this highest quality music in its full range of tone quality and volume. Therefore, the situation obviously requires connecting transmission paths so perfect in their characteristics that reproduction 100 or 200 miles away may not suffer in comparison with reproduction which may be only 100 or 200 feet from the source of music.

There are several respects in which a long line circuit possibly may distort the speech or music passed over it, unless considerable effort is expended to overcome these tendencies. For example, there may be frequency-amplitude distortion; i.e., all the notes and overtones may not be transmitted with the proper relative volumes. Similarly there may be phase or delay distortion, the different frequencies may not arrive at the receiving end of the line circuit in the same time relationships in which they originated. A line circuit is subject also to possible inductive disturbances from other communication circuits ("crosstalk"), or from power or miscellaneous circuits which cause "noise" at the receiving terminal. If the circuit contains amplifiers, transformers, and inductances having magnetic cores, it is subject to possible nonlinearity effects; i.e., the current at the receiving end of the line may not follow exactly the amplitude variations of the current applied to the transmitting end or, what is more important, spurious intermodulation frequencies may be generated within the transmission circuit and mar the purity of the musical tones. The problem of reproduction in auditory perspective, using two or three paralleling

* Fifth paper in the Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering*, January, 1934.

channels, also adds the requirement that these channels must be substantially identical in their transmission characteristics.

With the exception of the last, all these aspects of the problem are, of course, not peculiar to symphony music transmission. They exist as part of the problem of satisfactorily transmitting any telephone message. However, the requirements of this new high quality transmission have set a new high standard of refinement, even as compared with that required for ordinary radio chain broadcasting. For example, ordinary telephone message transmission commonly is carried out by circuits having a frequency range not exceeding 200 to 3000 cycles per second. Much present-day radio broadcasting involves a transmission band only from about 100 to 5000 c.p.s. This new high quality transmission, however, requires a range from approximately 40 to 15,000 c.p.s. Further, with reference to the required freedom from interference, ordinary radio broadcasting seldom exceeds a volume range greater than 30 decibels. The new high-quality system, however, requires a volume range of at least 65 db, which is more than 3,000,000 to 1 expressed as a power ratio.

In considering the specific problem of transmitting from Philadelphia to Washington for the demonstration given on April 27, 1933, several alternative methods of providing the required transmission paths presented themselves. The arrangement chosen consisted in bridging the distance between the two cities by means of carrier channels over cable conductors. From the telephone toll office in Philadelphia to the toll office in Washington, three carrier transmission paths were provided in which the music frequencies were stepped up from their normal position in the audible range to considerably higher frequencies. The frequency range from 40 to 15,000 c.p.s. picked up by the microphones was transmitted over line circuits in a range from 25,000 to 40,000 c.p.s. After being thus stepped up in frequency, the high frequency currents were applied to three non-loaded pairs in an all-underground cable which was equipped with repeaters at approximately 25-mile intervals. At Washington, step-down or demodulation apparatus restored the frequencies to their normal position in the spectrum.

For transmission between the auditorium in Philadelphia and the toll office there, a distance of approximately three miles, and for transmission in Washington between the telephone toll office and the auditorium, about half this distance, normal frequency transmission over small-gauge pairs in ordinary exchange cables was employed.

The use of the carrier method for the long distance transmission has several advantages. In general, it permits multiplex operation; i.e., more than one message or program on the same pair of wires. As a matter of expediency in this particular case this feature of operation was not used, and three separate pairs were employed, one for each channel. In the future the same technique undoubtedly would permit two or possibly more of these extra-broad-band transmission paths to be obtained on the same pair of conductors. The most important reason for choosing the carrier method rather than transmission in the natural audio-frequency range in this particular case was that, because all other transmission circuits in the same cable were at a considerably lower frequency and because the lead sheath of the cable acts efficiently at the high frequencies to shield the pairs from induced disturbances from the outside, it offered a special freedom from crosstalk and noise.

With these arrangements, which will be described in somewhat greater detail in what follows, requirements of transmission were met very satisfactorily and the reproduction of the symphony music in Washington with the orchestra playing in Philadelphia suffered not the least in comparison with the reproduction of the same program in an auditorium in Philadelphia located but a few feet from the hall in which the orchestra played. It is believed that, if necessary, by the use of the same principles, line circuits may be set up and comparable quality reproduction given throughout the country. However, as will be evident from part of the discussion which follows, in some respects the problem of meeting the requirements in transmission between Philadelphia and Washington was not as difficult as might be encountered in other localities. Hence even more complex arrangements might be necessary if it were desired to establish such transmission circuits to other points, and particularly for greater distances.

Line Circuits

There are several all-underground cables between Philadelphia and Washington. As described in a paper [1] by Clark and Green given before the A. I. E. E. in 1930, recently laid cables contain several 16-gauge conductors distributed throughout the cross section of the cable for possible use as program circuits in chain broadcasting. These pairs, however, ordinarily are loaded and equipped with repeaters at approximately 50-mile intervals so that they transmit a frequency range up to about 8000 c.p.s.

In one of the cables several pairs of this type had not yet been loaded, and these pairs were used for this newer transmission. Because of the higher frequencies employed and the greater attenuation encountered, it was necessary to install repeaters at more frequent intervals. As may be noted in Fig. 1, the normal cable layout between Philadelphia and Washington includes two intermediate repeater

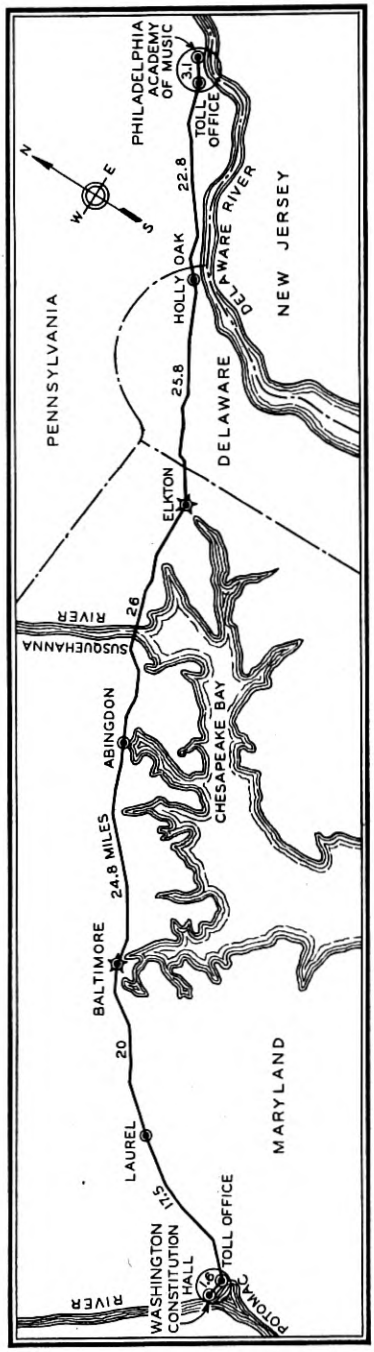

Fig. 1—Geographical layout of 3-circuit communication line used to carry Philadelphia symphony music to Washington, D. C., for reproduction in auditory perspective.

stations, one at Elkton and one at Baltimore. Additional repeater stations were established accordingly at in-between points—Holly Oak, Abingdon, and Laurel. One of these repeater points, Holly Oak, was established in a local telephone office. No such convenient housing existed at the other two points, and it was necessary to establish new repeater stations. These were small metal structures large enough to house only the repeaters, their power supply, and testing equipment. This apparatus was arranged to be normally non-attended, various switching actions being remotely controlled from the nearest regular repeater station.

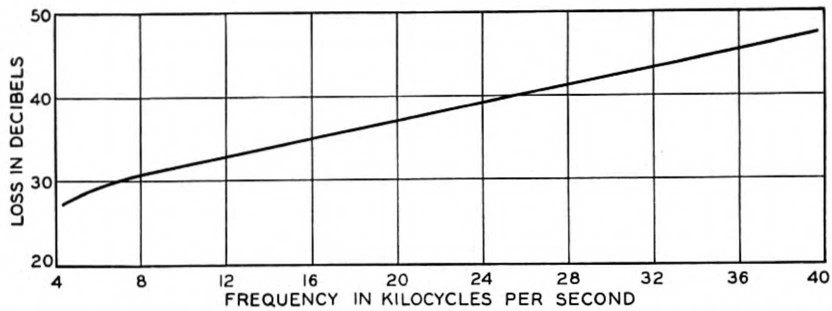

Fig. 2—Line attenuation characteristic of typical repeater section.

The line attenuation between repeater points is shown in Fig. 2. It may be noted that the attenuation is approximately 50 db for the highest carrier frequency involved. A diagram showing the variations in power level as the carrier waves traverse the complete circuit is shown in Fig. 3. Because of the variation in attenuation over the

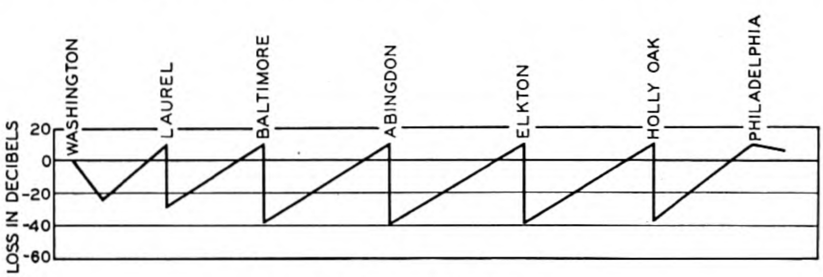

Fig. 3—Transmission level diagram.

frequency range employed it was necessary, of course, to use equalizers at the input of each repeater; i.e., networks having an attenuation variation with frequency approximately the inverse of that of the line circuit.

Noise

In setting up these circuits various tests, including measurements of noise currents picked up by the conductors to be employed, were made prior to the actual installation of the apparatus. It was discovered that on the cable circuits north of Baltimore these pairs were picking up sufficient noise even at the higher frequencies to constitute a possible limitation in the volume range that might be delivered. This noise was generated chiefly as a by-product of relay and other similar operations within the Baltimore office and was propagated over the longitudinal circuits of various pairs in the cable from which, by induction, it entered the special selected pairs. As a remedial measure, longitudinally acting choke coils applied to all but the specially selected pairs in the cable greatly reduced the noise. Shielding and physical separation were employed in the Baltimore office to prevent induction between the repeaters and the connection to the main cable. If it is desired to use existing cables for carrier transmission, particularly for such high grade transmission circuits, it seems likely that filtering arrangements of this kind, or other precautions, generally will be required.

Carrier Apparatus

The carrier system employed may be characterized briefly as single-sideband carrier-suppressed, with perfectly synchronized carrier frequencies of 40,000 c.p.s. Most present-day commercial telephone carrier systems are of the single-sideband carrier-suppressed type. Suppressing one sideband saves frequency space and suppressing the carrier reduces the load on the line amplifiers or repeaters. Ordinarily the exact synchronization of the carrier frequencies at the sending and receiving ends is not required for message telephone service.

Obtaining a single sideband after modulation commonly is carried out by providing band filters which transmit the desired sideband and suppress the unwanted sideband. For the requirements of message telephone transmission this does not impose severe requirements in the design of filters because audio frequencies less than about 100–200 c.p.s. ordinarily are not transmitted, in which case, if the filter in suppressing the unwanted sideband tends to cut off the lower frequencies of the desired sideband, it is not important.

For the requirements of this new high quality system, however, where it was desired to transmit all frequencies to at least as low as 40 c.p.s., the problem was considerably more difficult. Two alternatives presented themselves in the design of the required filters. The first consisted in attempting to provide the required selectivity in the filters themselves, perhaps supplementing the actions of inductance coils and condensers (which normally make up such a filter structure) by quartz crystals to provide the sharp selectivity required on the sides of the band. The other alternative consisted in providing a filter of moderate selectivity so that in the neighborhood of the carrier frequency the unwanted sideband is not completely suppressed, and in arranging that the resultant reproduced music at the receiving terminal is obtained by the proper coordination of the desired and the vestige of the unwanted sideband. The "vestigial" sideband method was decided upon. Although this does not require filters having particularly sharp selectivity on the sides of the band, it does, however, impose more severe requirements upon the control of the phase characteristics of the filters in the neighborhood of the carrier frequency. It makes it necessary also to have the carrier frequencies at the sending and receiving ends not only synchronized, but phase controlled as described later.

For the modulating elements in the system at both the sending and receiving terminals, copper oxide rectifying disks were chosen. These elements can be made very simple. In stability, with respect to transmission loss and the ability to suppress the unwanted carrier frequency by balanced circuits, this arrangement is superior to the usual vacuum tube circuits.

In Fig. 4 is shown schematically the arrangements of the carrier circuit at the transmitting and receiving ends. At the transmitting terminal the circuit from the microphones is led first through low- and high-pass filters to limit the bands to the desired width; i.e., 40 c.p.s. to 15,000 c.p.s. The 40-cycle limiting filter was included because tests had demonstrated that lower frequencies are not required for the satisfactory transmission of music of symphony character, and because it was feared that occasional high energy pulses of subaudible frequency might cause overloading. When these 40-cycle filters are omitted, as was done in tests, the carrier channels are capable of transmitting frequencies down to and including zero frequency, a characteristic which could not possibly be obtained in a single sideband system by other than such a vestigial sideband technique.

As may be noted further in Fig. 4, carrier current is supplied to the rectifying disks of the modulator along with the incoming music frequencies. The balancing connection of the four rectifying disks making up the modulator is arranged to suppress the carrier frequency, the final degree of suppression being adjusted by means of the variable condenser and resistance shown, which were included to make up for slight dissimilarities in the characteristics of the individual copper disks. A very high degree of carrier suppression can be achieved by

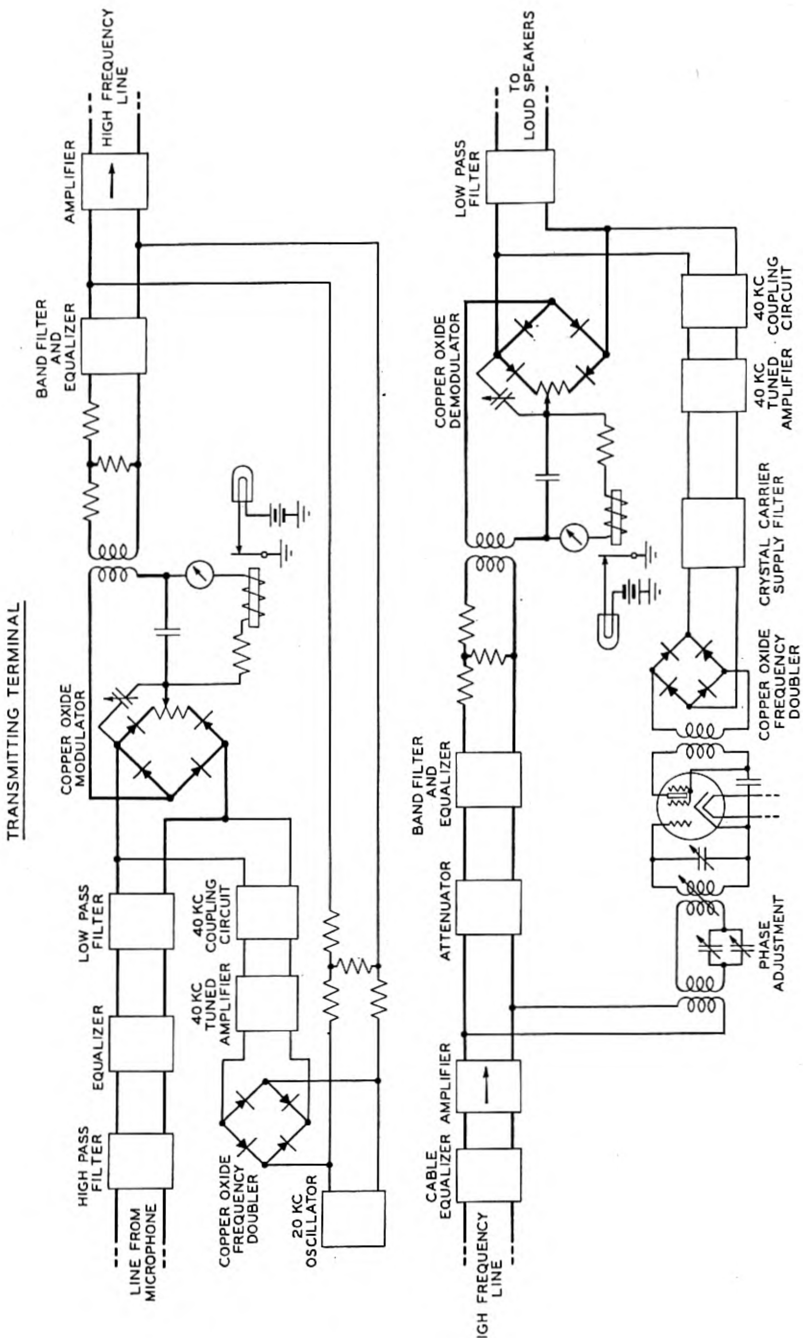

Fig. 4—Schematic diagram of carrier terminal circuits.

this means. No difficulty was experienced in maintaining a ratio of at least 60 db between the carrier voltage applied to the unit and the residual carrier current not completely balanced out. Over short periods an even higher degree of balance can be readily obtained.

There is a certain amount of electrical noise generated in the rectifying disks over and above that caused by thermal agitation [2, 3] effects. The amount of this noise compared with the maximum permissible modulation output determines the volume range possibilities of a modulator of this type. Measurements indicated that this range was approximately 90 db, which obviously was more than sufficient to meet the requirements desired.

The circuit includes a relay and a meter through both of which flows the d.-c. component produced by the rectification of the carrier frequency. These supplementary units give a check on the magnitude of the carrier supply and afford an alarm in case of failure. From the modulator unit the circuit is connected to the band filter which transmits only the lower sideband lying between approximately 25,000 and 40,000 c.p.s. and the vestige of the upper sideband. From the band filter the currents are led to an amplifier and thence to the line circuit leading to the farther terminal.

It may be noted that at the transmitting terminal the 40,000-cycle carrier current is derived from a 20,000-cycle oscillator by passing its output through a series of copper oxide rectifiers connected to form a frequency doubler. Part of the originally generated 20,000 cycles also is connected to the input of the transmitting amplifier and sent over the line to be used in producing the 40,000-cycle carrier supply for demodulation.

At the receiving terminal a similar modulation or demodulation process occurs through the use of copper oxide disk circuits. A relay and meter also are included in the circuit to check the carrier supply, in this case providing also a check or pilot of the transmission over the long line circuit. The 20,000-cycle synchronizing current is selected at the receiving terminal, amplified and applied to a frequency doubler, and thence applied to the demodulator circuit. The input of this carrier supply circuit includes also a phase adjusting variable condenser arrangement so that the phase of the carrier supplied to the demodulator may be adjusted properly in relation to that of the carrier supplied to the modulator at the sending end. An interesting feature of the receiving terminal carrier supply is the quartz crystal filter employed to select the 40,000-cycle carrier after frequency doubling. The transmission characteristic of this extremely selective filter is shown in Fig. 5.

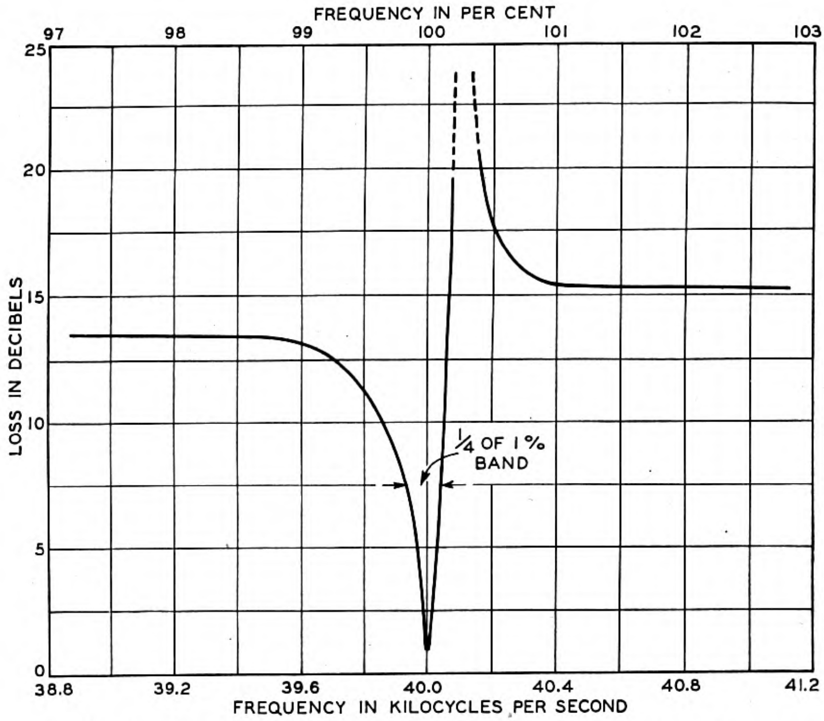

Fig. 5—Transmission characteristic of carrier supply crystal filter.

Filters

The transmission characteristics of the carrier channels are determined largely by the filters and associated equalizers. The filters principally affecting transmission are the band filters. Identical units are employed at the sending and receiving ends. The transmission and phase shift characteristics of one of these units are shown in Fig. 6. These band filters are equalized to produce the desired squared band characteristic.

The characteristics in the frequency region near the carrier (i.e., at 40,000 cycles) are shown on a large scale. This region is of particular interest because it is here that the degree of success in the application of the vestigial sideband method, for the purpose of insuring the satisfactory transmission of the low music frequencies, is determined. If for a given frequency interval above the carrier the phase change is arranged to be equal and opposite to that of the same frequency interval below the carrier, then in the action of demodulation the demodulated current produced by the action of one sideband adds itself arithmetically to that produced by the other sideband. It will be noted that this desirable phase characteristic has been achieved closely in the characteristics shown. If, in addition, the attenuation

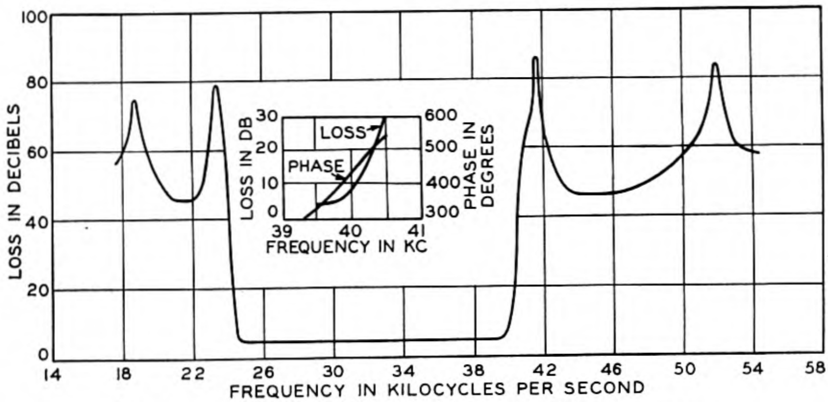

Fig. 6—Transmission and phase characteristics of band filter.

loss in the filter is adjusted so that the sum of the regular and vestigial sideband amplitudes corresponding to the low music frequencies is substantially constant and equal to the amplitude of the frequencies at midband, the desired flat transmission characteristic is assured.

As was noted previously, this action can be carried out only if the phase angle of the receiving carrier is properly related to that of the sideband frequencies and, in turn, to the carrier applied to the modulator at the transmitting terminal. The curves shown in Fig. 7

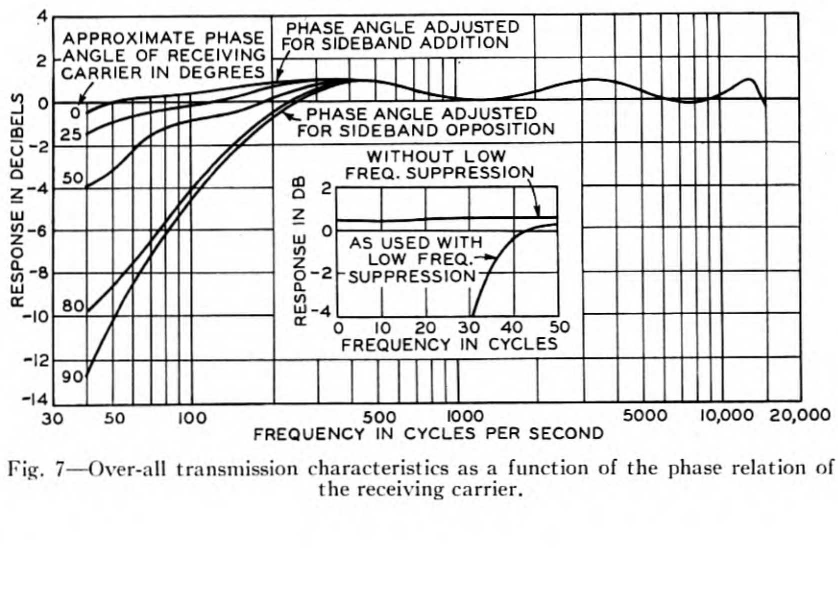

Fig. 7—Over-all transmission characteristics as a function of the phase relation of the receiving carrier.

illustrate the influence that the phase adjustment of the carrier frequency has on the transmission of the lower frequencies in a system of this kind.

The upper curve shows the transmission frequency characteristic of one of the carrier channels measured from terminal to terminal between distortionless lines, when the phase angle of the receiving carrier is adjusted for its optimum value. Under these conditions the vestigial sideband and normal sideband supplement each other in their effects to produce substantially flat transmission. (The insert indicates the sustained transmission toward zero frequency when the 40-cycle highpass filter is omitted from the circuit.) It may be noted also that with this proper phase adjustment the full band transmission characteristic provided is substantially flat within a fraction of a decibel from 40 c.p.s. to 15,000 c.p.s. The lower curves indicate successively what happens if the phase angle of the receiving carrier is adjusted different amounts from the optimum adjustment. It may be noted that for a 90-degree departure the transmission of a 40-cycle tone over the carrier channel would suffer more than 12 db in comparison with a 1000-cycle tone.

Repeaters

As noted previously, the line circuit between Philadelphia and Washington included five intermediate repeater points. A schematic drawing of the apparatus installed at each point is shown in Fig. 8.

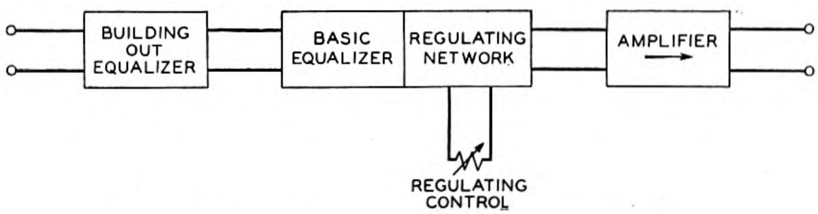

Fig. 8—Schematic diagram of repeater station apparatus.

The amplifiers at these points, as well as those used at the transmitting and receiving terminal, consisted of a new form of amplifier employing the principle of negative feed-back. The principal virtues of amplifiers of this type are their remarkable stability with battery and tube variations and great freedom from nonlinearity or modulation effects. Each amplifier is supplemented at its input by an equalizer designed to have its attenuation approximately complementary in loss to that of the line circuit in a single section. The amplifiers actually employed for the purpose were taken from a trial of a cable carrier system described in a recent A. I. E. E. paper by A. B. Clark and B. W. Kendall.[4]

The losses in the cable circuits do not, of course, remain absolutely constant with time, and slow variations due to change of temperature are compensated for by occasional adjustments of the variable equalizer arrangements provided. These adjustments were required only infrequently; approximately at weekly intervals because in an underground cable the temperature experiences only slow, seasonal variations.

As noted, new repeater stations were established at two points. The housing arrangements for one of these points, Abingdon, is shown in Fig. 9. The equipment at this repeater point also included relays remotely controlled from the nearest attended repeater station to permit the repeaters to be turned on and off at will and the power supply, which consisted of storage batteries, to be switched from the regular to the reserve battery or either battery put on charge if required.

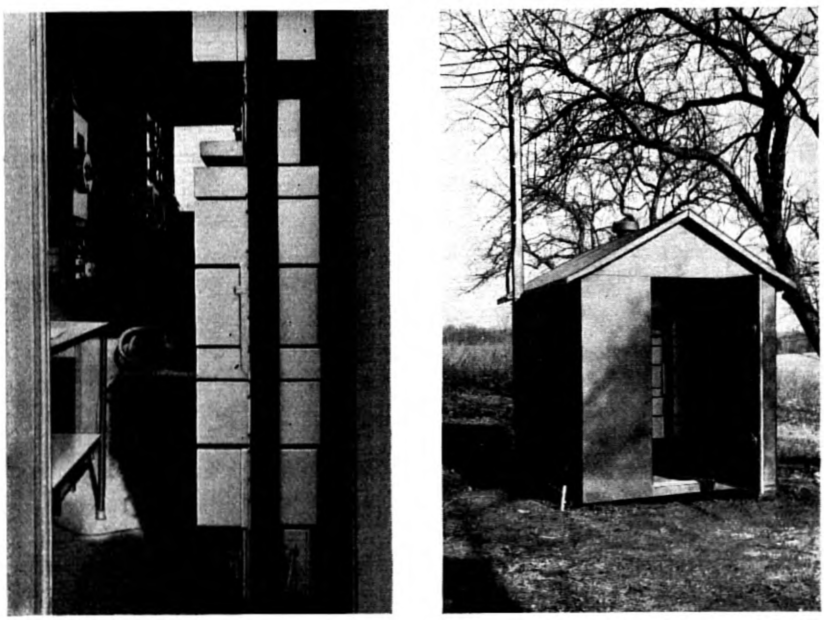

Fig. 9—Interior and exterior of special intermediate repeater station at Abingdon, Md.

Over-All Performance

While the system was set up specifically to provide transmission for the demonstration into Washington on April 27, 1933, it was operated over a period of several weeks and complete tests and measurements were carried out for the purpose of gathering information on cable carrier systems. The complete layout of apparatus and lines provided between Philadelphia and Washington is shown in Fig. 10.

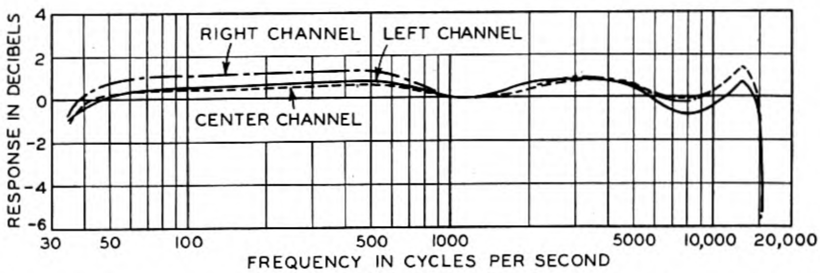

Fig. 11 (right)—Frequency characteristics over carrier channels used between Philadelphia and Washington, D. C.

The over-all frequency transmission characteristics of the three channels that were set up are shown in Fig. 11. These curves differ from those shown in Fig. 7, and include the complete high frequency line circuit with its 150 miles of cable, repeaters, equalizers, and other equipment. It may be seen that between the desired frequency limits the circuit is substantially flat in transmission performance to within $\pm$ 1 db. Various noise measurements made on the over-all circuit indicated that the circuits fully met the requirements that had been set up, and that the line and apparatus noise was inaudible in the auditorium at Washington even during the weakest music passages. The circuit also was found to be free from nonlinear distortion to a satisfactory degree. Harmonic components generated when single-frequency tones were applied to the channels at high volumes were found with one unimportant exception to be more than 40 db below the fundamental.

As a means of obtaining a further increase in volume range, which was not actually required for this demonstration, tests were made with a so-called predistortion-restoring technique. In this the higher frequency components of the music were transmitted over the carrier channels at a volume much higher than normal in relation to the volume of the lower frequencies. By this means any noise entering the carrier channels at frequencies equivalent to the higher music frequencies is greatly minimized in effect.

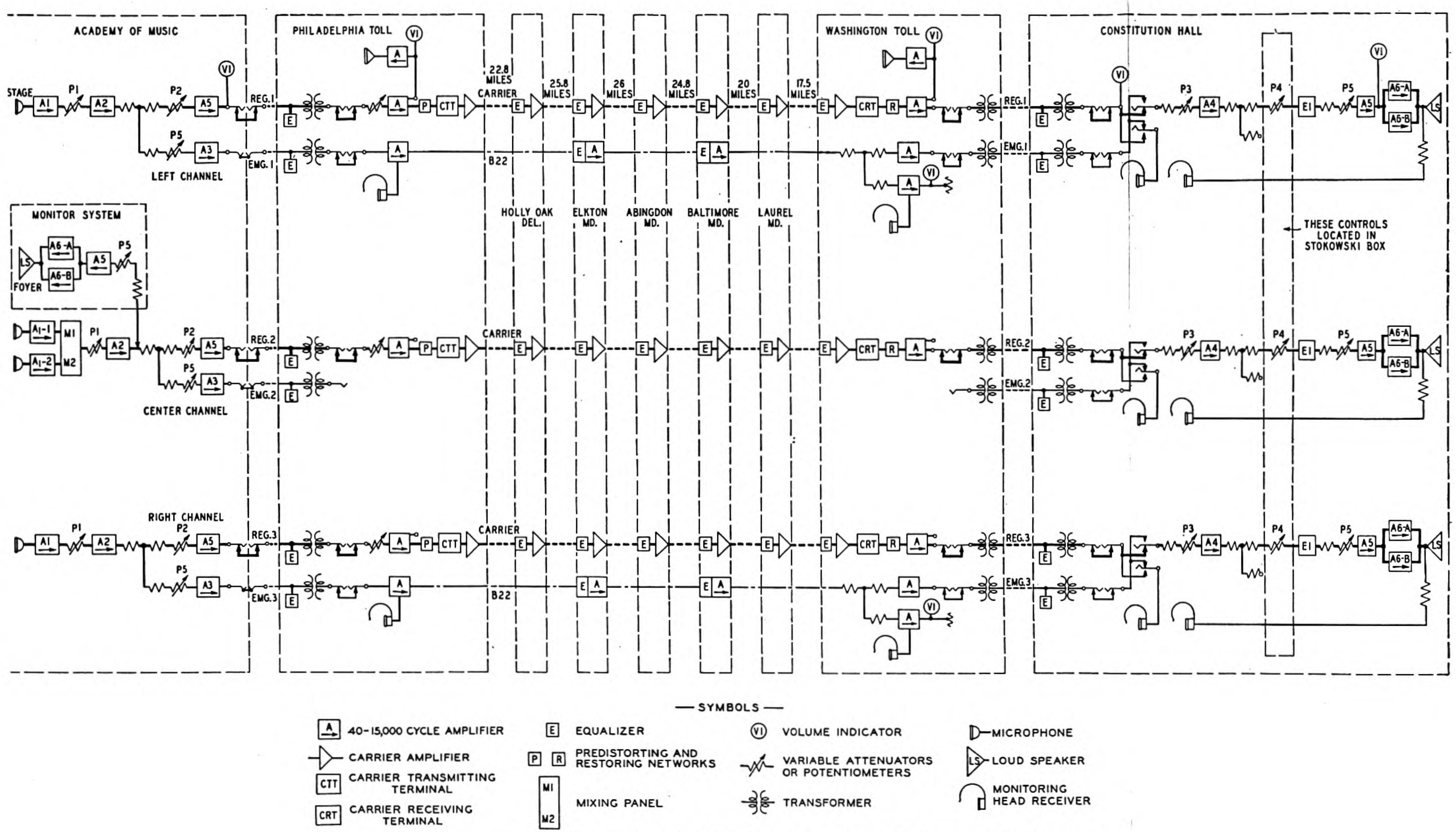

Fig. 10—Schematic diagram of circuit layout for 15-kc channel used for symphonic program demonstration.

This predistortion is accomplished by including in the circuit at the input to the modulator a network having relatively high loss for the lower frequencies and tapering to low loss for the higher frequencies. Its maximum loss is compensated for by adding in the circuit an equivalent amount of additional amplification. The characteristics of such a network are illustrated in Fig. 12. To restore the normal volume relationships between the different tones and overtones a

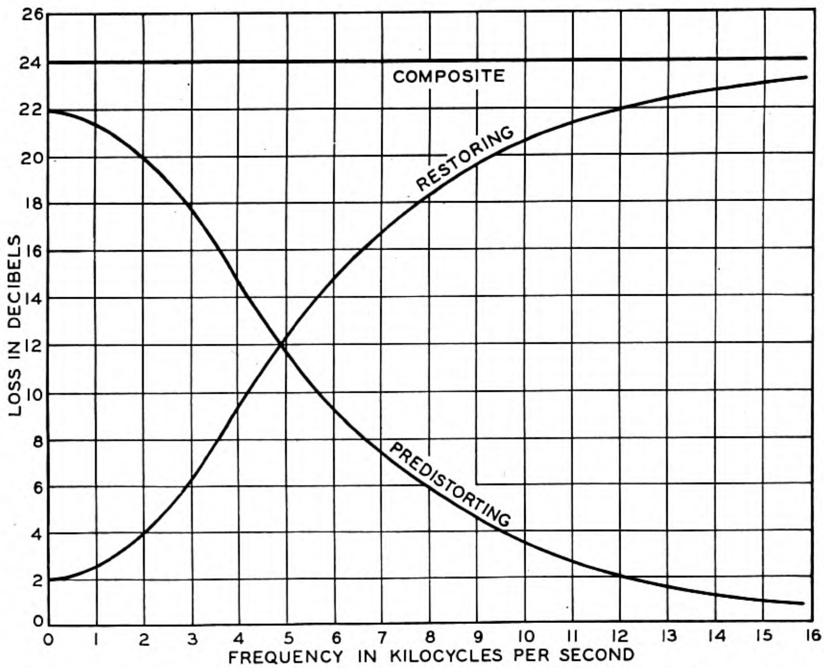

Fig. 12—Attenuation characteristics of "predistorting and restoring" networks.

restoring network having complementary transmission frequency characteristics is, of course, included at the output of the receiving circuit. It was found with this predistortion-restoring technique that a volume range increase of something like 10 db could be obtained over the circuits described.

There is available also another method which might have been employed for obtaining a further increase in volume range. This method, the so-called volume compression-expansion system, very likely will be necessary if in the future it is desired to obtain such high quality circuits on long routes where the carrier frequency range is being used also for regular telephone message transmission or for other purposes, and where the problem of freedom from noise and crosstalk no doubt will be more serious than experienced in the Philadelphia-Washington demonstration. Such a volume compression-expansion system requires additional apparatus at the sending and receiving terminals of the line circuit. At the sending end this apparatus is used to raise in volume the weak passages of the music or other program for transmission over the line circuits in order that the proper ratio between the desired program and unwanted noises may be retained. At the receiving terminal coordinating apparatus reexpands the compressed volume range to the volume range originally applied to the transmitting terminal.

In the demonstration, to provide supplementary control features required by Dr. Stokowski at Washington for communicating with the orchestra at Philadelphia, additional wire circuits were established between these points. Order wire circuits also were provided for communication between the terminals and repeater points to make possible the location troubles if any should arise. Rather elaborate switching means were included at the terminals to permit switching the carrier channels to different microphones and to different amplifier equipment at the loud speaker end. To take care of the contingency of a cable pair failure, spare pairs of wires were made available to be switched in at short notice. Fortunately, none of the reserve facilities actually were required for the demonstration.

References

1. "Long Distance Cable Circuits for Program Transmission," A. B. Clark and C. W. Green. *A. I. E. E. Trans.*, v. 49, 1930, p. 1514–23.
2. "Thermal Agitation of Electricity in Conductors," J. B. Johnson. *Phys. Rev.*, v. 32, 1928, p. 97.
3. "Thermal Agitation of Electric Charge in Conductors," H. Nyquist. *Phys. Rev.*, v. 32, 1928, p. 110.
4. "Carrier in Cable," A. B. Clark and B. W. Kendall. *Elec. Engg.*, July 1933, p. 477–81.

System Adaptation*

By E. H. BEDELL and IDEN KERNEY

A communication system for the pick-up and reproduction in auditory perspective of symphonic music must be designed properly with respect to the acoustics of the pick-up auditorium and the concert hall involved. The reverberation times and sound distribution in the two auditoriums, the location of the microphones and loud speakers, and the response-frequency calibration of the system and its equalization are considered. These and other important factors entering into the problem are treated in this paper.

WHEN the effect of music or the intelligibility of speech is spoiled by bad acoustics in an auditorium, the audience is well aware that acoustics do play a most important part in the appreciation of the program. One may not be conscious of this fact when the acoustical conditions are good, but a simple illustration will show that the effect still is present. Thus, of the sound energy reaching a member of the audience as much as 90 per cent may have been reflected one or more times from the various surfaces of the room, and only 10 per cent received directly from the source of the sound.

In listening to reproduced sound in an auditorium or concert hall, the effect of the room acoustics is perhaps even more important, for in this case the audience does not see any one on the stage and must rely entirely upon the auditory effect to create the illusion of the presence there of an individual or a group. Imperfections in the reproduced sound that are caused by defects in the acoustics of the auditorium may destroy the illusion and be ascribed improperly to the reproducing system itself.

In some types of reproduced sound, radio broadcast for example, where the reproduction normally takes place in a small room, the attempt is made to create the illusion that the listener is present at the source.[1, 2] In the case considered here, however, where symphonic music is reproduced in a large auditorium, the ideal is to create the illusion that the orchestra is present in the auditorium with the audience. Since the orchestra is playing in one large room and the music is heard in another, the acoustical conditions prevailing in both must be considered.

* Sixth and final paper in the Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering*, January, 1934.

Pick-Up Conditions

The source room is the auditorium of the American Academy of Music in Philadelphia. This room has a volume of approximately 700,000 cubic feet, and a seating capacity of 3000. Measured reverberation time curves for this auditorium, and preferred values [3, 4] for a room of this volume, are given in Fig. 1. It may be seen that with a

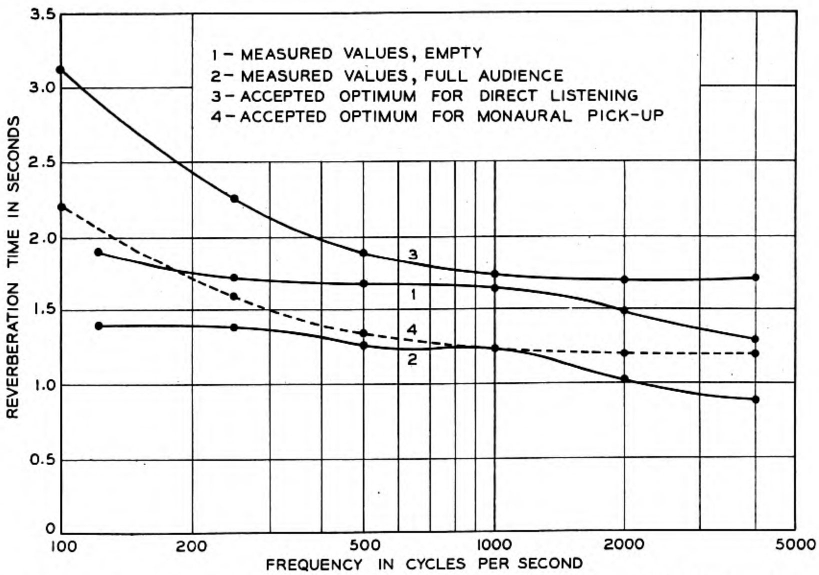

Fig. 1—Reverberation characteristics of Academy of Music, Philadelphia, Pa.

full audience this room might be considered somewhat dead, but would be considered generally satisfactory for pick-up either with or without an audience. A floor plan of the Academy auditorium and stage, showing the location of the three microphones used, is given in Fig. 2. The microphone positions were selected after judgment tests using several locations and are much nearer the orchestra than they would be for single channel pick-up.[2] The use of the microphones near the orchestra results in picking up a high ratio of direct to reverberant sound and thus reduces the effect of reverberation in the source room upon the reproduced music. A high ratio of direct sound is desirable in the present case also because of the use of three channels. The perspective effect obtained with three channels depends to a considerable extent upon the relative loudness at the three microphones, and since the change in loudness with increasing distance from the source is marked for the direct sound only, and not for the reverberant, there would be a definite loss in perspective effect if the microphones were placed at a greater distance from the orchestra. This effect is discussed more fully in another paper of this symposium.

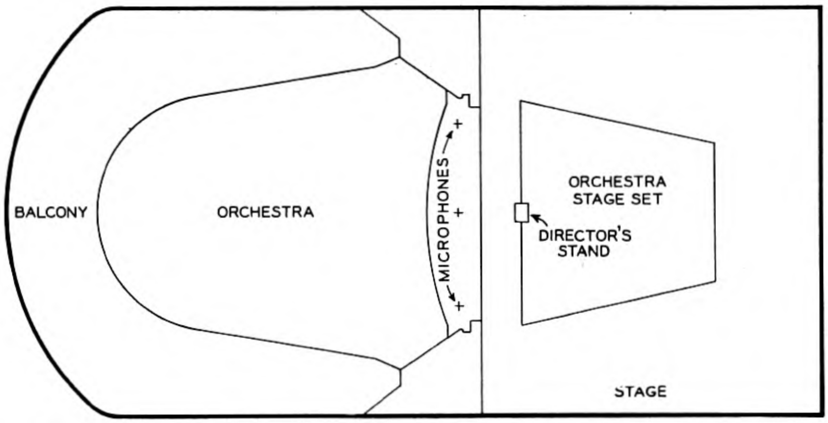

Fig. 2—Floor plan of Academy of Music, showing location of microphones.

With the microphones located close to the orchestra their response-frequency characteristics will be essentially those given by the normal field calibration, since relatively little energy is received from the sides and back. For a distant microphone position it would be necessary to use the random incidence response characteristic, which differs from the normal because of the variation in directional selectivity of the microphones as the frequency varies. This difference in response characteristic depends upon the size of the microphone and may amount to as much as 10 db at 10,000 c.p.s. It may be pointed out here that this difference in response is one factor frequently overlooked in the placement of microphones.

In addition to the three microphones regularly used, a fourth was provided to pick up the voice when a soloist accompanied the orchestra. In this case only the two side channels were used for the orchestra, the voice being transmitted and reproduced over the center channel. The solo microphone was so shielded by a directional baffle that it responded mainly to energy received from a rather small, solid angle. This arrangement permitted independent volume and quality control for the vocal and orchestral music.

The Concert Hall

The music was reproduced before the audience in Constitution Hall in Washington, D. C. This hall has a volume of nearly 1,000,000 cubic feet, and a seating capacity of about 4000. A floor plan of the auditorium showing the location of the loud speakers and of the control equipment is given in Fig. 3. The loud speakers are placed so that each of the three sets radiates into a solid angle including as nearly

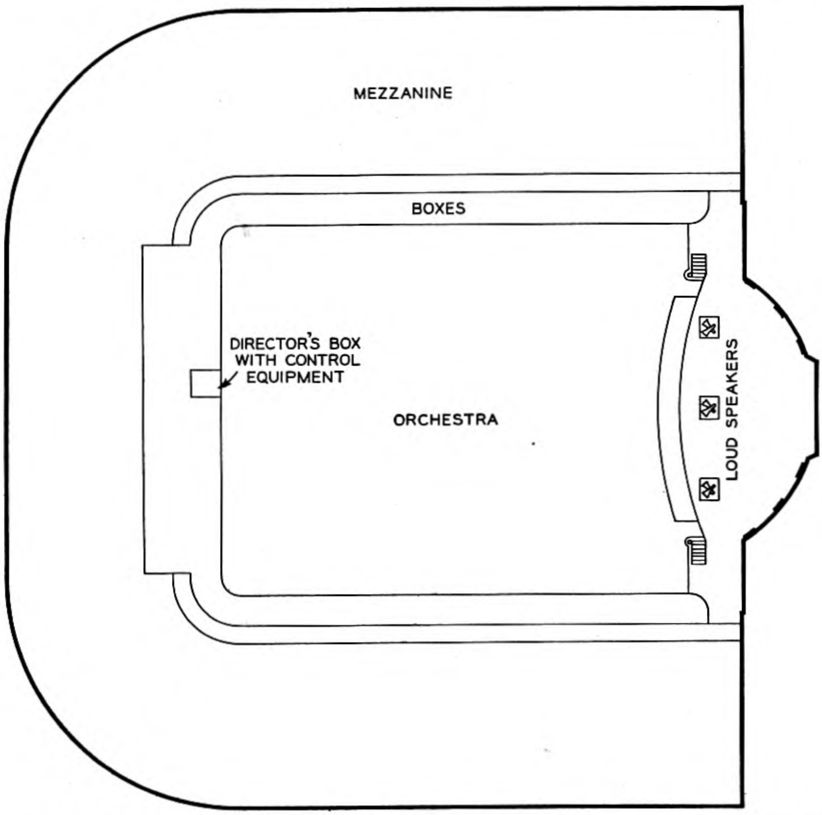

Fig. 3—Floor plan of Constitution Hall, Washington, D. C., showing locations of loud speakers.

as possible all the seats of the auditorium. Figure 4 shows the reverberation-frequency characteristics of Constitution Hall. The values given by the curve for the empty hall were measured through the use of the three regular loud speakers and several microphone positions in the room. The values for the hall with an audience present were calculated from known absorption data for an audience, and the optimum values are taken from accepted data for an auditorium of the volume of this one.[3] The reverberation times were considered satisfactory and no attempt was made to change them for this demonstration. The reverberation time measurements for both Constitution Hall and the Academy of Music were made with the high speed level recorder.[5] This instrument measures and plots on a moving paper chart a curve

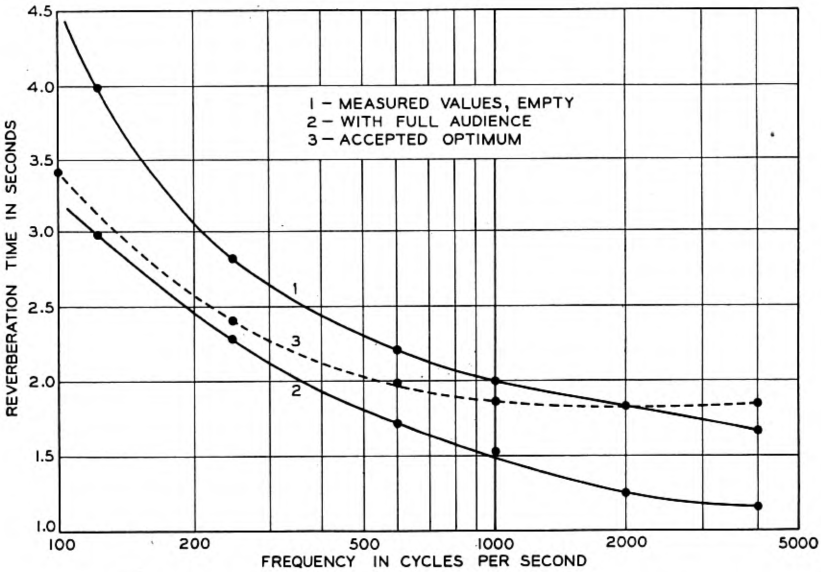

Fig. 4—Reverberation characteristics of Constitution Hall.

the ordinate of which is proportional to the logarithm of the electrical input furnished to it. When used in connection with a microphone for reverberation time measurements, curves are obtained showing the intensity of sound at the microphone during the period of sound decay. The rates of decay, and hence the reverberation times, are obtainable immediately from the slopes of these recorded curves and the speed of the paper chart.

Calibration of the System

In calibrating the system, a heterodyne oscillator connected to the loud speakers through the amplifiers was used. The oscillator was equipped with a motor drive to change the frequency, and as the frequency was varied through the range from 35 to 15,000 c.p.s. the sound was picked up with a microphone connected to the level recorder. Continuous curves of microphone response as a function of frequency thus were obtained for several positions in the auditorium, and for each channel independently. These response curves provided a check on a uniform coverage of the audience by each loud speaker, and also provided data for the design of the equalizing networks required to give an over-all flat response-frequency characteristic. If the system, including the air path from the loud speakers to one position in the auditorium, is made flat, it will not, in general, be flat for other positions or for other paths in the room. This variation in characteristic is due partly to the variation in the ratio of direct to reverberant sound, and partly to the fact that the sounds of higher frequency are absorbed more rapidly by the air during transmission.[6, 7] This latter effect is of considerable importance; it depends upon the humidity and temperature of the air, and may cause a loss of more than 10 db in the high frequencies at the more distant positions in a large auditorium. Some compromise in the amount of equalization employed therefore is necessary. Probably the most straightforward procedure would be to design the networks according to the response curves obtained with the microphones near the loud speakers. This would insure that for both the response measurements and the pick-up the microphone characteristics would be the same, and any deviation from a uniform response in the microphones would be corrected for in this way, along with variations in the loud speaker output. This procedure was modified somewhat for the case under discussion, however, because by far the greater portion of the audience was at a distance from the stage such that they received a relatively large ratio of reverberant sound, and it was believed that a better effect would be achieved by equalizing the system characteristic in accordance with response measurements taken at some distance from the loud speakers.

Control Equipment

In addition to the equalizing circuits used to obtain a uniform response characteristic, two sets of quality control networks which could be switched in or out of the three channels simultaneously were employed. One set modified the low frequencies as shown at A, B, and C of Fig. 5, while the other gave high frequency characteristics as shown at D, E, F, and G. These latter networks permitted the director to take advantage of the fact that the electrical transmission and reproduction of music permits the introduction of control of volume and quality which can be superimposed on the orchestral variations. Quality of sound can be divorced from loudness to a greater degree than is possible in the actual playing of instruments, and the quality can be varied while the loudness range is increased or decreased. Electrical transmission therefore not only enlarges the audience of the orchestra, but also enlarges the *capacity* of the orchestra for creating musical effects.

The quality control networks and their associated switches were mounted in a cabinet (Fig. 6) at the right side of the director's position. Continuously variable volume controls for the three channels were mounted on a common shaft and housed in the center cabinet of Fig. 6.

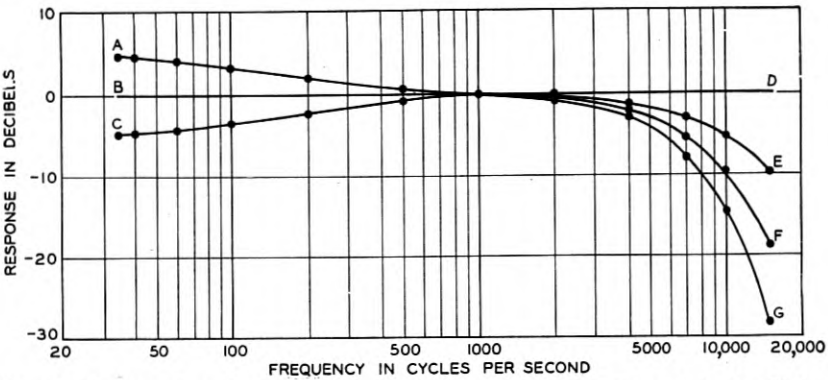

Fig. 5—Transmission characteristics of quality control networks used in the Philadelphia-Washington experiment.

A separate control for the center channel was provided when that was used for the soloist. In addition to the high quality channels certain auxiliary circuits were supplied to aid the smoothness of performance. Supplementing the order wire connecting all technical operators, a monitor circuit was provided in the reverse direction. The microphone was located on the cabinet before the director, and loud speakers were connected in the control rooms and on the stage with the orchestra, enabling the control operator to hear what went on in the auditorium and allowing the director to speak to the orchestra. Two useful signal circuits were employed; one giving the orchestra a "play" or "listen" signal, and at the same time connecting either the auditorium or the orchestra's loud speakers, respectively; the other being a

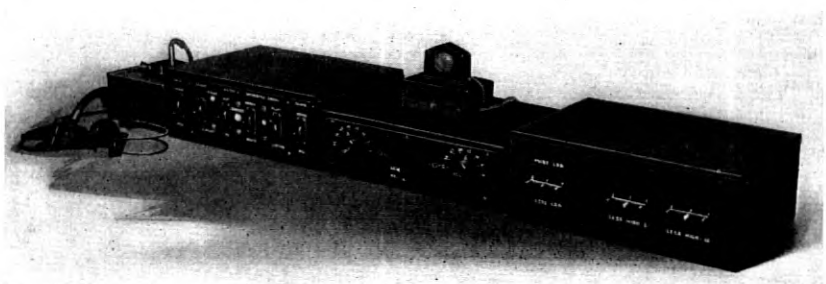

Fig. 6—Cabinets housing quality control networks and providing communication facilities for operation.

"tempo" signal to the assistant director leading the orchestra that could be operated during the rendition of the music. The switches for the auxiliary circuits and the order wire subset are shown at the control operator's position at the left in Fig. 6.

That a reproducing system may have quite different characteristics in different auditoriums is well illustrated in the case of the two halls considered here. From Fig. 3 it may be seen that in Constitution Hall the stage is built into the auditorium itself, and that there is no back stage space. The Academy of Music, however, has a large volume back stage. When the orchestra plays in the Academy the reflecting shell shown in Fig. 2 is used to concentrate the radiated sound energy toward the audience. When the reproducing system was set up in the Academy the shell could not be used because of the stage and lighting effects desired, and a large part of the energy radiated by the loud speakers at the low frequencies was lost back stage. The loss of low frequency energy is attributable partly to the fact that the loud speakers cannot well be made as directional for the very low frequencies as for the higher. The loss amounts to about 10 db at 35 c.p.s., and becomes inappreciable at 300 c.p.s. or more, as measured in comparable locations in the two auditoriums. This difference in characteristics emphasizes the fact that for perfect reproduction the acoustics of the auditorium must be considered as a part of the system, and that in general the equalizing networks must have different characteristics for different auditoriums.

References

1. "Acoustics of Broadcasting and Recording Studios," G. T. Stanton and F. C. Schmid. *Jour. Acous. Soc. Am.*, v. 4, No. 1, part 1, July 1932, p. 44.
2. "Acoustic Pick-Up for Philadelphia Orchestra Broadcasts," J. P. Maxfield. *Jour. Acous. Soc. Am.*, v. 4, No. 2, Oct. 1932, p. 122.
3. "Optimum Reverberation Time for Auditoriums," W. C. MacNair. *Jour. Acous. Soc. Am.*, v. 1, No. 2, part 1, Jan. 1930, p. 242.
4. "Acoustic Control of Recording for Talking Motion Pictures," J. P. Maxfield. *Jour. S. M. P. E.*, v. 14, No. 1, Jan. 1930, p. 85.
5. "A High Speed Level Recorder for Acoustic Measurements," E. C. Wente, E. H. Bedell, K. D. Swartzel, Jr. Unpublished paper presented before the Acous. Soc. Am., May 1, 1933.
6. "The Effect of Humidity Upon the Absorption of Sound in a Room, and a Determination of the Coefficients of Absorption of Sound in Air, V. O. Knudson. *Jour. Acous. Soc. Am.*, v. 3, No. 1, July 1931, p. 126.
7. "Absorption of Sound in Air, in Oxygen, and in Nitrogen—Effects of Humidity and Temperature," V. O. Knudson. *Jour. Acous. Soc. Am.*, v. 5, No. 2, Oct. 1933, p. 112.

Abstracts of Technical Articles from Bell System Sources

Effects of Rectifiers on System Wave Shape.[1] P. W. BLYE and H. E. KENT. Operation of mercury arc rectifiers generally results in increased harmonic currents in the rectifier supply circuits and may result in increased harmonic voltages. While these harmonics usually are not serious from the standpoint of the power system, they may result in interference to communication circuits exposed to the power circuits. This paper presents a method of computing these harmonic voltages and currents, and discusses methods of coordinating telephone systems and a-c. power systems supplying rectifiers.

Joint Use of Poles with 6,900-Volt Lines.[2] W. R. BULLARD and D. H. KEYES. A plan has been developed for joint occupancy of poles by power and telephone circuits in the Staten Island, N. Y. area, involving 6,900-volt distribution. The aim of this plan is to secure to the public and to the power and telephone companies over-all safety, convenience, and economy. Results of this cooperative study of joint use are presented in this paper.

Sound Film Printing—II.[3] J. CRABTREE. The production of sound-film prints from variable density negatives by the Model D Bell & Howell printer has been studied from the point of view of high-frequency response and uniformity of product. The account of this study, begun in Part I, is continued here, with particular reference to the degree of influence of slippage on the high-frequency response, occasioned particularly by non-conformity of the perforation pitch of the negative and positive films. It is found that to improve printing conditions in practice, it is first necessary to achieve consistency in the pitch of the processed negative and positive materials and to make the pitch of the processed negative 0.0004 inch less than that of the positive raw stock.

The Determination of the Direction of Arrival of Short Radio Waves.[4] H. T. FRISS, C. B. FELDMAN, and W. M. SHARPLESS. In this paper are described methods and technique of measuring the direction with

[1] *Elec. Engg.*, January, 1934.
[2] *Elec. Engg.*, December, 1933.
[3] *Jour. S. M. P. E.*, February, 1934.
[4] *Proc. I. R. E.*, January, 1934.

which short waves arrive at a receiving site. Data on transatlantic stations are presented to illustrate the use of the methods. The methods described include those in which the phase difference between two points constitutes the criterion of direction, and those in which the difference in output of two antennas having contrasting directional patterns determines the direction. The methods are discussed first as applied to the measurement of a single plane wave. Application to the general case in which several fading waves of different directions occur then follows and the difficulties attending this case are discussed.

Measurements made with equipment responsive to either the horizontal or the vertical component of electric field are found to agree.

The transmission of short pulses instead of a steady carrier wave is discussed as a means of resolving the composite wave into components separated in time. More detailed and significant information can be obtained by this resolving method. The use of pulses indicates that (1) the direction of arrival of the components does not change rapidly, and (2) the components of greater delay arrive at the higher angle above the horizontal. The components are confined mainly to the plane of the great circle path containing the transmitting and receiving stations.

A method is described in which the angular spread occupied by the several component waves may be measured without the use of pulses.

Application of highly directional receiving antennas to the problem of improving the quality of radiotelephone circuits is discussed.

Electron Diffraction and the Imperfection of Crystal Surfaces.[5] L. H. Germer. Bragg reflections are obtained by scattering fast electrons (0.05A) from the etched surfaces of metallic single crystals. The surfaces studied are a (100) face of an iron crystal, (111) face of a nickel crystal and (110) face of a tungsten crystal. In each case the reflections occur accurately at the calculated Bragg positions with no displacement due to refraction. A given reflection is found, however, even when the glancing angle of the primary beam differs considerably from the calculated Bragg value—by over 1.0° in some cases—so that several Bragg orders occur simultaneously. The accuracy with which this glancing angle must be adjusted is a measure of the degree of imperfection of the crystal. From the electron experiments, estimates are made of the widths at half maximum of electron rocking curves. These widths are 0.8° for the iron crystal, 1.5° for the nickel crystal and somewhat over 1.0° for the tungsten crystal. X-ray rocking curves for these same crystals are much narrower, although the observed

[5] *Phys. Rev.*, December 15, 1933.

widths vary considerably with the treatment of the surfaces. It is concluded that the values obtained from the electron measurements apply to projecting surface metal only, and that *the degree of misalignment is much greater at the surface than deep down within the crystal.* Furthermore, even the x-rays [Mo $K\alpha$ radiation $- 0.71$A] are not sufficiently penetrating to yield values certainly characteristic of these metal crystals.

Mutual Impedance of Grounded Wires Lying on the Surface of the Earth when the Conductivity Varies Exponentially with Depth.[6] MARION C. GRAY. This paper presents a formula for the mutual impedance of any insulated wires of negligible diameter lying on the surface of the earth and grounded at their end-points, on the assumption that the conductivity of the earth varies exponentially with depth. Various special cases are briefly discussed.

Signals and Speech in Electrical Communication.[7] JOHN MILLS. This book is written by a member of the technical staff of Bell Telephone Laboratories who is well-known for his text on "Radio Communication" (1917) and the more popular presentations of "Within the Atom" (1921) and "Letters of a Radio-Engineer to His Son" (1922). In this book he presents for the general reader a synthesis of the electrical arts of communication in terms of their general fundamental principles. In separate chapters, which are discrete essays in popular and semi-technical language, the fundamental principles of dial operation, transmitters and receivers, loading coils, repeaters, multi-channel or carrier systems, and transoceanic radio-telephony are graphically expounded. The entertaining treatment of engineering achievements in allied fields of the sound picture, broadcasting, television, stereophonic reproduction and the teletypewriter, will intrigue the layman and assist him in acquiring a general understanding of these highly technical developments.

Some Earth Potential Measurements Being Made in Connection with the International Polar Year.[8] G. C. SOUTHWORTH. For several years the Bell System has been studying the relation between radio transmission and earth potential disturbances. A paper dealing with this subject was published in 1931. Prompted by the needs of the International Polar Year, together with the prospect that further work would throw additional light on the nature of radio transmission, the work was extended somewhat in 1932.

[6] *Physics*, January, 1934.
[7] Published by Harcourt Brace and Company, New York, N. Y., 1934.
[8] *Proc. I. R. E.*, December, 1933.

It is expected that useful correlation will be found between the normal earth potential effects which occur day after day during undisturbed periods and the corresponding diurnal and seasonal variation of radio transmission. It seems entirely probable, for instance, that earth potentials are but the terrestrial manifestations of certain changes taking place in the Kennelly-Heaviside layer which may not be found by other methods.

This paper is intended to serve mainly as a progress report outlining briefly the methods and scope of the work and showing the type of data being obtained. It leaves to a later date most of their correlation and their interpretation. The data here presented are in a conventional form used by other investigators for many years. Their value lies mainly in their extent and in the rather wide range of circumstances under which they were obtained.

Investigation of Rail Impedances.[9] HOWARD M. TRUEBLOOD and GEORGE WASCHECK. Measurements of impedance made on five sizes of rails and on two types of bonds are reported in this paper; the investigation covered a range of current per rail of 20 to 900 amperes, and frequencies of 15 to 60 cycles per second. Results are given in a form convenient for engineering use, and include information for applying corrections for bond impedance and for temperature.

[9] *Elec. Engg.*, December, 1933.

Contributors to this Issue

H. A. Affel, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1914; Research Assistant in Electrical Engineering, 1914–16. Engineering Department and the Department of Development and Research, American Telephone and Telegraph Company, 1916–34. Bell Telephone Laboratories, 1934–. Mr. Affel has been engaged chiefly in development work connected with carrier telephone and telegraph systems.

E. H. Bedell, B.S., Drury College, 1924; University of Missouri, 1924–25. Bell Telephone Laboratories, Acoustical Research Department, 1925–. Mr. Bedell's work has had to do mainly with studies of sound absorption and transmission and allied subjects in the field of architectural acoustics.

Arthur G. Chapman, E.E., University of Minnesota, 1911. General Electric Company, 1911–13. American Telephone and Telegraph Company, Engineering Department, 1913–19, and Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. Mr. Chapman is in charge of a group engaged in developing methods for reducing crosstalk between communication circuits, both open wire and cable, and evaluating effects of crosstalk on telephone and other services.

R. W. Chesnut, A.B., Harvard University, 1917; War Department of French Government, 1917; U. S. Army, 1917–19. Engineering Department, Western Electric Company, 1920–25; Bell Telephone Laboratories, 1925–. Mr. Chesnut has been engaged in the development of carrier telephone and long-wave radio systems.

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.

Frederick S. Goucher, A.B., Acadia University, 1909; A.B., Yale University, 1911; M.A., Yale, 1912; Ph.D., Columbia University, 1917. Western Electric Company, 1917–18. University College, London, 1919. Research Laboratories, General Electric Company, Limited, North Wembley, England, 1919–26. Bell Telephone Laboratories, 1926–. Dr. Goucher has been engaged in a study of the carbon microphone.

Iden Kerney, B.S. in Communication Engineering, Harvard Engineering School, 1923. Development and Research Department, American Telephone and Telegraph Company, 1923–34. Bell Telephone Laboratories, 1934–.

R. H. Mills, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1916. Western Union Telegraph Company, 1916–18. Western Electric Company, Transmission Development Branch, 1918–25. Bell Telephone Laboratories, Apparatus Development Department, 1925–. Mr. Mills is responsible for the development of carrier frequency filters for use in commercial communication systems.

E. O. Scriven, B.S., Beloit College, 1906; Instructor, Fort Worth University, 1906–08; S.M., Massachusetts Institute of Technology, 1911. Engineering Department, Western Electric Company, 1911–25. Bell Telephone Laboratories, 1925–. Mr. Scriven is in charge of the electrical design of special products apparatus other than radio.

W. B. Snow, A.B., Stanford University, 1923; E.E., 1925. Engineering Department, Western Electric Company, 1923–24. Acoustical research, Bell Telephone Laboratories, 1925–. Mr. Snow has been engaged in articulation testing studies and investigations of speech and music quality.

J. C. Steinberg, B.Sc., M.Sc., Coe College, 1916, 1917. U. S. Air Service, 1917–19. Ph.D., Iowa University, 1922. Engineering Department, Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Dr. Steinberg's work since coming with the Bell System has related largely to speech and hearing.

A. L. Thuras, B.S., University of Minnesota, 1912; E.E., 1913. Laboratory assistant with U. S. Bureau of Standards, 1913–16. Graduate student in physics, Harvard, 1916–17. Bell Telephone Laboratories, 1920–. At the Laboratories, Mr. Thuras has worked on the study and development of electro-acoustic devices and instruments.

E. C. Wente, A.B., University of Michigan, 1911; S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1914; Ph.D., Yale University, 1918. Engineering Department, Western Electric Company, 1914–16 and 1918–24; Bell Telephone Laboratories, 1924–. As Acoustical Research Engineer, Dr. Wente has worked principally on general acoustic problems and on the development of special types of acoustic devices.