The Bell System Technical Journal

July, 1926

The Power of Fundamental Speech Sounds

By C. F. SACIA and C. J. BECK

SYNOPSIS: This paper describes the continuing work on speech power by means of oscillographic studies of vowels, semi-vowels and consonants. A previous paper considered the characteristics of a few individual sounds from the power standpoint, but the principal emphasis was placed upon speech as a whole. In this later analysis, sounds are considered individually on the basis of instantaneous and mean power. A practical application of the results is suggested.

CONTINUING the work done on speech power by means of power oscillograms,[1] we have made additional reductions in the data relative to the vowels, semi-vowels and consonants and have also prepared a smaller amount of data on the power of the semi-vowels and the consonants from the amplitude oscillograms.[2] This is a preliminary study of the subject, at least in so far as the latter two classes of sounds are concerned, for these records of speech sounds were made to show all sounds in their true relative value hence the consonant sounds, being greatly inferior to the vowels were measurable to a correspondingly smaller degree of accuracy. We have gathered such data as the existing records could yield before future plans are completed to make a more comprehensive study of consonants.

Stop consonants are not so well characterized by the power data as are other types. The unvoiced stop consonants have two properties: a puff whose main frequency component is of the order of 50 cycles with a few ripples of high frequency; and a modifying effect upon the beginning or end of the vowel which immediately precedes or succeeds it. Hence, such a consonant is more of a controlling factor and lacks the essential properties of a discrete sound. In giving the data on the puff where it is measurable, we separate the low and high frequency components. In the case of the voiced stop consonants the vocal cord vibrations give the consonant more character of its own.

MEAN POWER AND PEAK POWER

In the paper on speech power and energy, the "mean power," P_m, was derived (in the case of the vowel sounds) as the mean of the power taken throughout the interval of the vocal cycle. By the assumption of an appropriate arbitrary interval instead, say of the order of one

[1] B. S. T. J. Vol. IV No. 4. "Speech Power and Energy," by C. F. Sacia.
[2] B. S. T. J. Vol. IV No. 4. "Sounds of Speech," by I. B. Crandall.

one-hundredth of a second, the definition applies as well to consonant sounds and in addition has the same practical significance as that of the mean power of a vowel.

Mean power is thus a variable function of time, starting from zero, rising to a maximum and eventually falling to zero again as the sound is being uttered.[3] In studying an aggregate of speech sounds it is impracticable to have the final results in terms of these mean power curves; the most important discriminant of such a curve of any sound is its maximum ordinate, P_m. This value was used in the earlier study and has been given the name "syllabic power" when used in connection with the syllable as a whole. In the present case we shall abbreviate by simply calling it the "mean power of the sound." Similarly, when we are considering the consonant apart from the rest of the syllable we select the maximum value of P_m for that consonant.

Likewise, in considering the instantaneous power of a sound we select the height of the greatest peak occurring therein and for convenience we call it the "peak power."

All the averages hereinafter tabulated are the arithmetic averages of such maximum ordinates and not the integrated averages.

Normal and Conversational Values

We specify "normal" values as those derived from monosyllables spoken disconnectedly without accent but also without being slighted; while "conversational" values are derived from ordinary conversational speech. It does not follow that the arithmetic average of conversational values for a given sound should equal the average of the normal value, for the reason that some sounds are slighted much more frequently than others, as we shall see later.

The Consonants and Semi-Vowels

Of these sounds two independent sets of data are available: instantaneous peak power and mean power. The former is summarized in Table I. To explain the table in detail we take as an example the consonant, "t" as in "tap." There being one observation upon each of two speakers, the greatest observation showed 19 microwatts (peak) from the lips of the one speaker while the other speaker reached a peak of 13 microwatts, and the average of these two is 16. As in the paper on Speech Power and Energy, the corresponding values of power intensity in microwatts per square centimeter at the condenser transmitter are given in the group at the right. Since the relating factor is

[3] See "Speech Power and Energy," Fig. 1, page 628, for comparison of instantaneous and mean powers.

TABLE I

Normal Values of Peak Power in Microwatts for Two Speakers

(A) CONSONANTS

Consonant		Total from Voice			Per Cm² at Trans.		
Symbol	Key	Max.	Min.	Ave.	Max.	Min.	Ave.
b	bat	7	7	7	0.06	0.05	0.06
p	pot	7	6	6	0.06	0.05	0.05
*p	pot	128	0	64	1.04	0.	0.52
d	dot	7	1	4	0.06	0.01	0.04
t	tap	19	13	16	0.15	0.11	0.13
g	get	9	7	8	0.07	0.06	0.06
k	kit	9	4	6	0.07	0.03	0.05
dh	that	10	8	9	0.08	0.06	0.07
th	thin	1	0	1	0.01	0.	0.01
*th	thin	30	0	15	0.24	0	0.12
v	vat	29	21	25	0.23	0.17	0.20
*f	for	53	10	31	0.42	0.08	0.25
f	for	4	2	3	0.04	0.02	0.03
j	jot	26	23	24	0.21	0.19	0.20
ch	chat	61	43	52	0.49	0.35	0.42
zh	azure	53	23	38	0.43	0.19	0.31
sh	shot	133	97	115	1.08	0.79	0.93
z	zip	42	21	31	0.34	0.17	0.25
s	sit	54	8	31	0.43	0.06	0.25

* Low frequency puff.

(B) SEMI-VOWELS

Semi-Vowel		Total from Voice			Per Cm² at Trans.		
Symbol	Key	Max.	Min.	Ave.	Max.	Min.	Ave.
l	let	226	37	131	1.83	0.29	1.06
ng	ring	169	25	97	1.36	0.20	0.78
n	no	74	21	47	0.59	0.17	0.38
m	me	198	23	111	1.60	0.18	0.89

NOTE: For these two speakers, the peak power of the succeeding vowel was as follows:

	Total	Per Cm²
ū (tool)	206	1.7
á (tap)	860	6.8
ē (teem)	241	1.9

about 127, the intensities 0.15, 0.11 and 0.13 are the first three numbers respectively divided by 127.

These values were derived by measuring the amplitudes of the abovementioned oscillograms of the acoustic pressure. The maximum or peak amplitudes of the consonant and the succeeding vowel were first measured; the square of the ratio between these is the ratio of the corresponding peak powers. Now the approximate peak powers of these vowels for the two speakers were found (see note under Table I) from the power oscillograms used in our study of speech power. Hence from the product we derive the approximate peak power of the consonant (or semi-vowel). Direct measurement of peak power from the latter oscillograms was impracticable because of the low sensitivity of the instantaneous power recorder [4] and the before-mentioned fact that the power of the consonants and semi-vowels is low relative to that of the vowels.

Since frequencies of the order of 50 cycles are of negligible importance in speech, the 50-cycle puff has been separated from the other components in the case of the unvoiced stop consonants. This is justified by the fact that the utterances of such a sound by two speakers may seem exactly alike to the careful listener, whereas a large puff may be present in one case and none in the other.

The values thus far considered represent "normal" values in speech —not accented and yet not slighted.

TABLE II

Conversational Values of Mean Power in Microwatts for 16 Speakers

(A) CONSONANTS

Consonant		Speaker's Power		Number of Measurable Obser-vations	Per Cm² at Trans.	
Symbol	Key	Max.	Av.		Max.	Av.
d	dot	2.9	0.08	4	0.023	0.0006
t	tap	6.0	0.14	14	0.049	0.0012
k	kit	4.8	0.34	20	0.039	0.0027
v	vat	2.4	0.03	1	0.019	0.0002
f	for	3.6	0.08	1	0.029	0.0006
j	jot	3.6	0.47	8	0.029	0.0038
ch	chat	7.9	1.44	19	0.064	0.0116
sh	shot	6.0	1.83	9	0.049	0.0148
z	zip	7.2	0.72	31	0.058	0.0058
s	sit	8.7	0.94	115	0.070	0.0076

(B) SEMI-VOWELS

Semi-Vowel		Speaker's Power		Number of Measurable Obser-vations	Per Cm² at Trans.	
Symbol	Key	Max.	Av.		Max.	Av.
l	let	9.6	0.33	13	0.078	0.0026
ng	ring	3.6	0.35	2	0.029	0.0028
n	no	18.0	2.11	146	0.145	0.0170
m	me	16.8	1.85	31	0.136	0.0149

[4] In recording the power, separate vibrators had been used for instantaneous and mean powers.

Our measurements of mean power, on the other hand, were made from power records of conversational speech, with a greater variety of observations and speakers. Stress, therefore, plays an important part here.

In Table II is given a compact summary of the direct measurements made on the power oscillograms. Thus consider "d" as in "dot." 2.9 microwatts was the greatest observed value for any speaker, while the average of all observations (including accented and unaccented utterances) was but 0.08. Only four observations, however, were large enough to be measured. As before, we give the corresponding intensities in microwatts per square centimeter at the transmitter in the next two columns.

To show the occurrence of stress in the utterance of these sounds in ordinary speech, we give in Fig. 1 the stress frequency-distribution curves [5] of several oft-occurring sounds. These curves are derived in the same manner as were the syllabic stress curves in the study of speech power. They exhibit the marked degree in which the consonants differ in stress for ordinary speech. For example, among the consonant sounds, "t" and "sh" represent extreme types. The former is either slighted or strongly accented with but little intermediate gradation while the blunt characteristic of the latter indicates the most nearly uniform distribution of stress into all shades from zero to maximum. Similarly with the three semi-vowels shown, "l" and "m" are extreme types.

The Vowels

Some attention was given to vowel power in the other paper where under the heading of "Relative Power of Vowels" (on page 634) were charted what we have classified as normal values of mean power. These were derived from the mean power curves of disconnected monosyllables. Although they were charted separately for male and female voices, we shall not differentiate between the two in the following. In Tables III and IV are summarized the four sets of data based upon the speech from 16 voices. Here we see the influence of stress by comparing the conversational and normal values. This effect is noteworthy in the case of "o" (ton) "à" (tap) and "i" (tip) which average considerably less power in conversational speech than in normal syllables. Another point of interest is the comparison of peak and mean values. For example, in the normal data, the ratio of peak to mean (i.e. the

[5] The abscissa represents the relative number of observations ($s/\underline{s}$) whose relative power values exceed the magnitude of the ordinate, n, a numeric varying between zero and one.

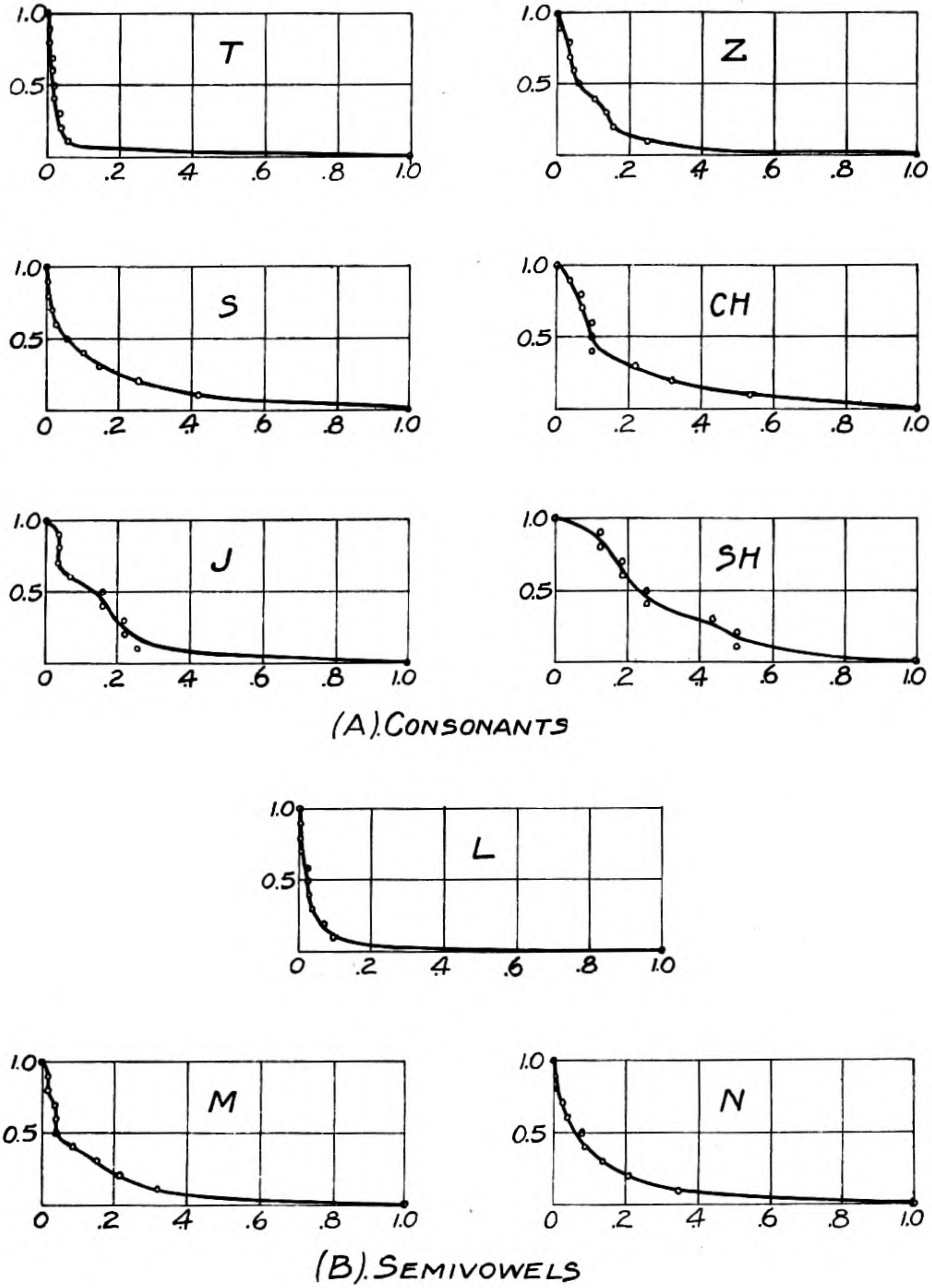


Fig. 1. Power Stress Curves.

TABLE III—VOWELS

Peak Power in Microwatts for 16 Speakers

Vowel		Total from Voice				Per Cm² at Trans.				Number of Measurable Observations Conversational Values
		Normal Values		Conversational Values		Normal Values		Conversational Values		
Symbol	Key	Max.	Av.	Max.	Av.	Max.	Av.	Max.	Av.	
ū	tool	620	290	760	180	5.0	2.3	6.1	1.5	61
u	took	890	470	—	—	7.2	3.8	—	—	0
ō	tone	1310	540	900	330	10.6	4.4	7.2	2.7	62
ò	talk	1240	630	1580	600	10.0	5.1	12.8	4.8	32
o	ton	1240	600	1720	300	10.0	4.8	13.9	2.4	248
a	top	1650	760	1580	660	13.3	6.1	12.8	5.3	127
à	tap	1860	1020	1380	290	15.0	8.2	11.1	2.3	38
e	ten	1720	660	1510	340	13.9	5.4	12.2	2.8	125
ā	tape	1380	580	1720	470	11.1	4.7	13.9	3.8	32
i	tip	1240	520	1330	190	10.0	4.2	10.8	1.5	198
ē	teem	1510	430	960	190	12.2	3.5	7.8	1.5	56
r̄	err	—	—	550	200	—	—	4.4	1.6	33

Note: The dash indicates that observations were not available.

TABLE IV—VOWELS

Mean Power in Microwatts for 16 Speakers

Vowel		Total from Voice				Per Cm^2 at Trans.				Number of Measurable Observations Conversational Values
		Normal Values		Conversational Values		Normal Values		Conversational Values		
Symbol	Key	Max.	Av.	Max.	Av.	Max.	Av.	Max.	Av.	
ū	tool	60	33	53	13	0.49	0.27	0.43	0.11	64
u	took	108	40	—	—	0.87	0.32	—	—	0
ō	tone	82	38	68	22	0.66	0.31	0.55	0.18	64
ò	talk	91	43	125	47	0.74	0.35	1.01	0.38	32
o	ton	84	33	107	15	0.68	0.27	0.86	0.13	284
a	top	111	48	130	34	0.89	0.39	1.05	0.28	128
à	tap	96	40	40	9	0.78	0.33	0.32	0.07	48
e	ten	79	27	88	17	0.64	0.22	0.71	0.13	141
ā	tape	58	26	62	20	0.47	0.21	0.49	0.16	32
i	tip	53	30	55	9	0.43	0.24	0.44	0.07	250
ē	teem	65	27	78	12	0.52	0.22	0.63	0.10	64
r̄	err	—	—	30	10	—	—	0.24	0.08	40

Note: The dash indicates that observations were not available.

square of the peak factor) is greater for centrally located vowels and is greatest for "à" (tap) as was mentioned in the earlier paper. Referring to the normal values of peak power we find a surprising degree of regularity in the increase of these values from a minimum for "ū" (tool) to a maximum for "à" (tap) and the falling off again to minimum for "ē" (teem). The one slight irregularity is the vowel "o" (ton). (We have omitted "r̄" (err) from this comparison because it has no well defined place on the Vietor triangle which forms the basis for this arrangement of the other vowels).

TABLE V—SPEECH SOUNDS

Speech Sound	Key	Relative Power, Arbitrary Units		C Relative Power Attenuation to give 80% Articulation
		A Mean Power Conversational values for 16 speakers	B Peak Power Normal values for 2 speakers	
ò	talk	1870	688	826
a	top	1380	1430	474
ō	tone	875	630	619
ā	tape	808	632	567
e	ten	664	975	364
o	ton	616	688	474
ū	tool	532	344	349
ē	teem	484	402	421
r̄	err	384	– see note	924
à	tap	366	2170	645
i	tip	346	688	295
n	no	84	78	36
m	me	74	185	38
sh	shot	73	192	216
ch	chat	58	87	64
s	sit	38	51	11
z	zip	29	52	17
j	jot	19	41	98
ng	ring	14	162	134
k	kit	14	10	43
l	let	13	218	157
t	tap	6	26	32
d	dot	3	7	60
f	for	3	6	9
v	vat	1	41	13
u	took	– see note	688	347
zh	azure	–	63	–
dh	that	–	15	–
g	get	–	13	60
b	bat	–	11	30
p	pot	–	11	24
th	thin	–	1	1

NOTE: The dash indicates that observations were not available.

Relative Power of Speech Sounds

A direct comparison of most of the fundamental sounds will now be made. In Table V—A are shown the conversational values (averaged) of the mean power for each sound for 16 speakers. The units are taken arbitrarily in order to show only the relative values. As might have been expected, the vowels rank the highest, the semi-vowels next and the consonants the lowest, although we find a few consonants interspersed among the semi-vowels. In Table V—B is the similar arrangement for the normal values of peak power for the two speakers. Data on a larger number of sounds are available for this group, but the same general order prevails: vowels, semi-vowels and consonants. Minor differences in order (note "v" as in "vat") may be expected to occur because of the influence of stress upon the conversational value. But in both cases the ratio of the maximum to the minimum is of the order of 2000. This similarity is striking in view of the difference in the modes of utterance and the numbers of speakers in the two cases.

Finally, in Table V—C are shown relative values[6] derived on the basis of relative attenuation in power required to bring the articulation (as judged by the average ear) to 80%. Since disconnected monosyllables

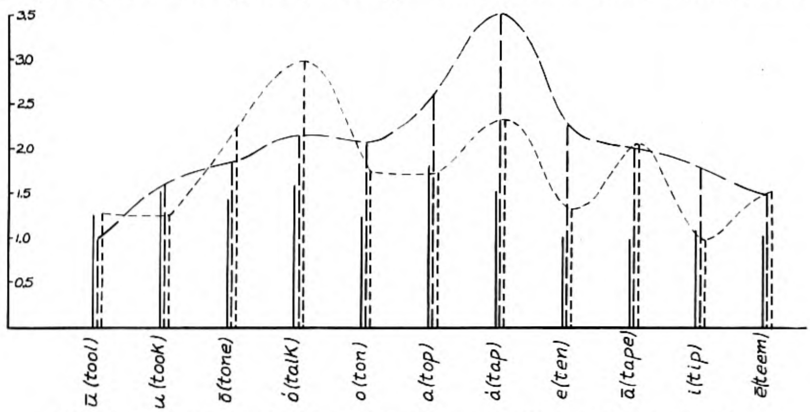


Fig 2. Comparative Chart Relative Normal Values of Vowel Sounds.
— — — — — —Peak Power.
————————Mean Power.
- - - - - - - - -Relative Power Attenuation Required to Give 80% Articulation.

were used in this test the values are normal values in our present category. Although the same general order of the other two tables

[6] Taken from the paper presented by Harvey Fletcher before the Modern Languages Association, December 1923. Values are there called relative "intensity" which term we avoid here because of the acoustic meaning already assigned to intensity: power per square centimeter.

prevails here, there are considerable differences throughout which may well be expected since the ear is used in making the balance. The frequency response characteristic of the ear is the complicating factor in this case. The ratio of maximum to minimum here is of the order of one thousand or about one-half the absolute power ratio found in the two preceding tables.

A more orderly comparison between power and "relative attenuation" exists in the case of the vowels alone as shown in the chart of Fig. 2. Thus the peak power and "relative attenuation" most nearly correspond at the ends of the chart (especially the left) where there is resonance of lower frequency in the vowels. The vowel "o" again shows a peculiarity in that the two trends—as shown by the envelopes—intersect here. Peak power predominates over "relative attenuation" in the three successive vowels "a," "à," "e," which have strong resonance in the region from 600 to 1200 cycles. The vowel "i" gives the only erratic turn in this comparison, differing considerably from the two adjacent vowels.

As for loudness in the ordinary sense, let us note a phenomenon of rather common occurrence in these days of good quality sound reproducing apparatus. One may be listening to well reproduced speech at ordinary volume when suddenly a slightly accented syllable containing "à" (tap) comes through with noticeable overload distortion and its accompanying disagreeable effect upon the ear. Although the listener does not judge this sound to be any louder than numerous accented sounds preceding and following it, still the fact remains that there has been considerable overload due to the peaks of the wave being cut off by the amplifier. Where do we look for the explanation? As noted in the earlier paper this vowel has the highest peak factor, and we have already seen in Table III that it normally contains the greatest peak power. In spite of this therefore, it would seem that the loudness of this sound does not predominate over the loudness of the sounds in the first half of the chart, as does the peak power. This phenomenon can also be demonstrated, for the vowel "ē" (teem) and to a lesser degree even for the vowels which intervene between these two in the tables and chart of the vowel sounds.

Extraneous Interference on Submarine Telegraph Cables

By J. J. GILBERT

SYNOPSIS: In order to avoid a considerable reduction in speed of operation, which would have resulted on account of the unusually large parasitic disturbances encountered in the neighborhood of New York, the New York-Azores permalloy loaded cable was equipped with a new type of earth connection consisting of a conductor extending 100 nautical miles to sea and there connected to earth through an artificial line.

This paper presents the theory of the new type of sea earthing arrangement and discusses the sources of extraneous interference and the manner in which it is picked up by submarine cables. A method is developed for estimating the magnitude of terminal extraneous interference in the case of any particular cable.

AMONG the factors limiting the speed of operation of long submarine telegraph cables one of the most important is the mutilation of the received signals by electrical disturbances picked up along the cable and transmitted with the incoming signal to the receiving instrument. The nature of this disturbance is shown in Fig. 1 which is an oscillographic record over a short period of time of the difference of potential across the terminals of the receiving instrument of a cable system, at a time when no signals were being received over the cable. Although the complete signal correction networks were not in circuit at the time this record was taken, the latter is representative of the form of the extraneous disturbance that would be superposed on the record of an incoming signal. It is evident that unless the signal amplitude is sufficiently large compared with the amplitude of interference, the latter will seriously interfere with the interpretation of the siphon recorder tape or with the functioning of relays operated by the signal current. That this condition constitutes a limit on the speed of operation of the cable is indicated by Fig. 2 which shows the amplitude of a signal, received over a typical transatlantic cable, as a function of

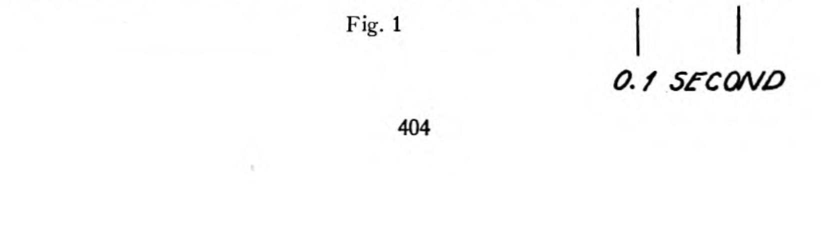
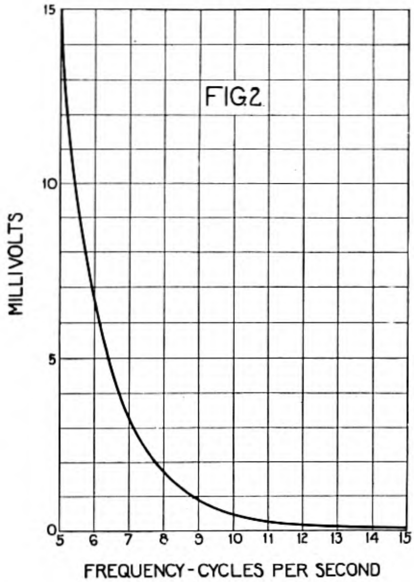


Fig. 1

the signal frequency.[1] It is evident, that corresponding to the minimum amplitude at which signals are just legible through interference, there is, for a given value of sending voltage, a maximum speed of signalling which cannot be exceeded, without danger of serious mutilation of the signal. If by any means the magnitude of the extraneous interference can be diminished, signals of smaller amplitude can be employed and the speed of operation consequently increased.



The present paper will be devoted to a description of the manner in which extraneous interference is picked up by submarine cables, with a discussion of the influence of various factors such as depth of water, cable structure and operating conditions. There will also be described a method of reducing interference by a modification of the cable structure. This method has been remarkably successful in the case of the New York-Azores continuously loaded cable,[2] and has helped to make available the great gain in operating speed due to continuous loading, which is the outstanding feature of this cable installation.

The disturbances encountered on submarine cables are due mainly to induction from extraneous electromagnetic fields in the sea water,

[1] The signal frequency is defined as the fundamental frequency involved in a succession of alternately positive and negative unit impulses.

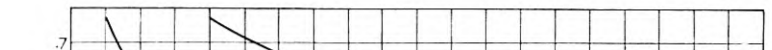
[2] Buckley, O. E., *Journal A. I. E. E.*, Vol. XLIV, p. 821, August 1925, *Bell System Technical Journal*, Vol. IV, No. 3, July 1925.

arising from a variety of sources, which may be broadly grouped into two classes. The first class comprises artificial sources, such as electrical power or railway systems in the neighborhood of the cable terminals. Currents circulating between the various earth connections of such systems give rise to electromagnetic fields in the earth and sea water, which fields may have the form of transient surges or pulses, or may be periodic in nature. The second class includes the various manifestations in the atmosphere or at the surface of the earth, such as electric or magnetic storms, which are also responsible for the disturbances in radio communication known as "static." Very little definite data is available regarding the magnitude and character of the natural disturbances affecting submarine cables, but it is found that, as in the case of static, the intensity of such effects is influenced by a number of factors, the season of the year and the geographical location being among the most important. At times of unusual activity, such as that accompanying the aurora polaris or local electrical storms, the voltages induced in the cable conductor are so large as to prohibit operation of the cable.

Except in the case where the source is in the immediate vicinity of the cable, the effect of any disturbance upon the cable can be considered as the result of a fluctuation of potential at the surface of a massive conducting medium, the ocean, which gives rise to electromagnetic waves which are propagated in all directions from the source and which penetrate the interior of the conducting medium according to the well-known laws governing "skin effect." Due to the presence of varying electric and magnetic fields in the sea water adjacent to the cable, an electromotive force is induced in each section of the cable conductor, and the resulting current is transmitted along the conductor to the cable terminal, combining with the currents due to electromotive forces induced in other sections to make up the total extraneous interference.

At the surface of the ocean the disturbance may take a variety of forms, for instance a succession of pulses or a train of damped oscillations. In any case the most convenient method of following the disturbance through the sea water into the cable conductor and along the conductor to the cable terminal is to consider the disturbance made up of a number of sinusoidal components of all frequencies from zero to infinity, the relative amplitudes and phases of the various components being determinable from the wave shape of the disturbance by the methods of Fourier analysis. The transmission characteristics of the interference transmission system at any particular frequency can then easily be studied, and finally the total effect of the original disturbance can be obtained by summation of disturbances of all frequencies.

The extent to which electrical disturbances penetrate below the surface of the ocean can be determined from the theory of induction of currents in continuous media, where it is shown that the components of the electric (E) and magnetic fields (H) parallel to and at a distance



z below the surface of an infinite plane conductor are given by the formulas:[3]

$$E = E_o\, \epsilon^{-kz},\ H = H_o\, \epsilon^{-kz},\ k = 2\pi\sqrt{2\lambda i f}, \tag{1}$$

where E_o and H_o are the values of E and H at the surface, λ is the electrical conductivity of the medium and f is the frequency. Employing the value of λ for sea water and expressing z in feet, gives

$$k = 1.35 \times 10^{-3}\sqrt{f}\ (1+i).$$

The curves of Fig. 3, computed from formula (1), indicate the manner in which sinusoidal disturbances of frequencies in the telegraph range are attenuated by various depths of sea water. It can be seen that the magnitude of a disturbance falls off rapidly as it penetrates the water; also that this attenuating effect is greater the higher the frequency. At a depth of one or two miles, at which the greater part

[3] Jeans "Electricity and Magnetism," 2nd Edition, p. 477.

of the typical transoceanic cable is submerged, only the extremely low frequency components of the surface disturbance are encountered to an appreciable degree. In the vicinity of the terminals, however, where the water is comparatively shallow, the cable is exposed to the higher frequency components of the disturbances, and it is usually in these sections that the greater part of the most troublesome disturbances is picked up. This is especially true in localities where the zone of shallow water extends a considerable distance from shore. Such a case is shown in Fig. 4, which represents a typical profile of the ocean bottom for the shallow water portion of a cable terminating at New York.

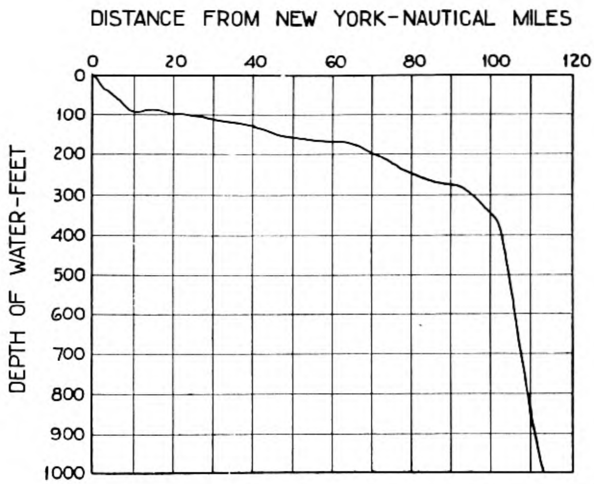


FIG. 4

The phenomena attending the induction of an electromotive force in the cable conductor by an electromagnetic field are rather complicated and difficult of exact computation. In the first place, on account of the change in electrical constants in passing from sea water to ocean bottom, the electric and magnetic field intensities in the neighborhood of the cable are somewhat different than indicated by equation (1). The influence of this factor upon the final result is in general small compared with that of the other factors that we are considering, and, on account of our lack of knowledge concerning the electrical characteristics of the ocean bottom, theoretical discussion would be of little practical value. A second factor is the shielding effect of the armor wires and metallic tapes surrounding the core. No attempt will be made in the present paper to work out an analytical solution of this problem. There is available, however, from a recent study of the problem of the sea return resistance of a submarine cable,[4] information that enables us to compare the behavior of various cable structures from the point of view of shielding. One of the results of this work was the determination of the degree to which the shielding effect of the metallic sheath around the cable causes the returning signal current to flow in this sheath rather than in the surrounding sea water. It is obvious that the greater the tendency of the metallic sheath to confine the return current to itself, the more effective the sheath will be in reducing the pick-up of interference. Allowing for the two effects just discussed, it is evident that the electromotive force induced in unit length of the cable conductor is given by an expression of the form

$$e = A\ E_o\, \epsilon^{-kz} \tag{2}$$

where A is a multiplier, the value of which will be determined only on a relative scale.

The electromotive force induced in any section of the cable conductor gives rise to sinusoidal currents and potentials which are transmitted in both directions along the conductor in accordance with well-known laws. For simplicity we will assume that the cable is terminated at both ends in its characteristic impedance, Z, the result corresponding to any other values of terminal impedances being readily determinable if needed.[5] Then an electromotive force $e\,dx$, induced in a short section of cable of length dx, distant x from the terminal, will result in a current

$$\frac{e\,dx}{2Z}\,\epsilon^{-\gamma x} \tag{3}$$

at the terminal. If the electromotive force per unit length e is picked up uniformly over a length of cable extending from $x=a$ to $x=s$, then since the impedance in each direction from the point is Z, the resulting current at $x=o$ will be

$$\frac{e}{2Z}\int_a^{a+s} dx\,\epsilon^{-\gamma x} \tag{4}$$

$$= \frac{e}{2Z}\,\epsilon^{-a\gamma}\,\frac{1-\epsilon^{-s\gamma}}{\gamma}$$

[4] Carson and Gilbert "Transmission Characteristics of a Submarine Cable," *Jour. Franklin Inst.*, Vol. 192, p. 705, 1921, and *Electrician*, Vol. 88, p. 499, 1922; *Bell System Technical Journal*, Vol. I, No. 1, July 1922.

[5] Heaviside, "Electromagnetic Theory," Vol. 2, p. 75.

Thus the effect at $x=o$ is the same as if an electromotive force $e\frac{1-\epsilon^{-s\gamma}}{\gamma}$ had been impressed at $x=a$.[6]

It would now be possible to assume a definite form of disturbance at the surface of the ocean, and by applying the principles that have been discussed in the preceding pages, to work out for any particular cable the wave shape of the resulting interference at the cable terminals. On account of our lack of knowledge as to what might be considered a typical disturbance at the surface of the ocean, such results would be merely speculative, and would be of no practical value in predicting the actual terminal interference that might be expected. A much better scheme is to compute for each cable, what may be called the interference susceptibility, this being defined, for a particular frequency, as the integral

$$\int A \,.\, \epsilon^{-kx} \cdot \epsilon^{-\gamma x} \cdot dx. \tag{5}$$

the integration extending over the entire cable. A is a factor which takes account of the shielding by armor wires, and changes at each point on the cable where the armoring changes. z is the depth of immersion at a distance x from the terminal, the relation between z and x being obtainable from the profile curve of the cable route. By comparing the susceptibility-frequency curves for two cables we can obtain an idea of the relative disturbances to be expected on the cables, with the possible exception of that part arising from sources in close proximity to the cables. For the latter type of interference special considerations are necessary.

In drawing conclusions from a susceptibility-frequency curve it is essential to bear in mind that, although the disturbance at the cable terminal is a composite of sinusoidal voltages and currents of all frequencies from zero to infinity, we are principally concerned with the

[6] An interesting conclusion to be drawn from equation (4) is that the contributions from various portions of a long section of cable due to a uniform disturbance tend to neutralize each other, on account of the fact that they arrive at $x=a$ in various phases. Since γ is equal to $\alpha+j\beta$, where α is the attenuation constant and β the phase constant, both per unit length, the quantity $\epsilon^{-s\gamma}$ can be represented graphically by a vector of length $\epsilon^{-s\alpha}$ and angle $(-s\beta)$. If α were zero the value of the factor $1-\epsilon^{-s\gamma}$ would be zero for $s\beta=o$, 2π, 4π, 6π, etc. That is the disturbance picked up over a length of cable $s=\frac{2n}{\beta}$, n being any integer, would have no effect at the terminal of the cable. On account of the fact that a is not zero, the quantity $\epsilon^{-s\gamma}$ is less than unity for all the above values of s except $s=o$, and complete neutralization of the disturbance does not occur. In the case of an inductively loaded cable, however, for a given value of α, β is many times greater than the value for the corresponding non-loaded cable. This means that neutralization of interference picked up on the loaded cable is much more complete than in the case of a non-loaded cable.

components lying within a certain frequency range, the limits of which depend upon the speed of signalling. This is due to the fact that the characteristics of an ordinary submarine cable are such that the low frequency components of a signal are transmitted with much less di-diminution of amplitude than are the higher frequency components. Consequently [7] it is found necessary, in order to render the signal intelligible, to employ a correcting network at the receiving terminal, one function of which is to attenuate the arriving low frequency components so that they finally are in the proper proportion to the higher frequency components. Also it is found that frequencies which are higher than about one and one-half times the signal frequency are not required in order to obtain intelligible signals, so that the receiving network can be designed to remove disturbances of the higher frequencies. The receiving apparatus therefore acts as a band filter towards the interference arriving at the terminal and emphasizes the part played by the components of interference of frequencies in the neighborhood of the signal frequency. On this account it is possible, in the majority of cases, to obtain the significant portion of the susceptibility-frequency curve by limiting the integration in (5) to the portion of the cable submerged to a depth of approximately 1000 feet or less, since, as has been previously indicated, only disturbances of extremely low frequencies are picked up on the deep water portion of the cable.

Given the problem of predetermining the interference at the terminal of a projected cable, the following procedure can be employed:

1. Over a period of time sufficiently long, a series of records of interference is taken on a cable terminating in the same general neighborhood as the proposed cable. Oscillographic records of the type shown in Fig. 1 are very desirable for this purpose.

2. From these records, and from the computed susceptibility-frequency curves of the existing and projected cables the interference on the latter can be predicted.

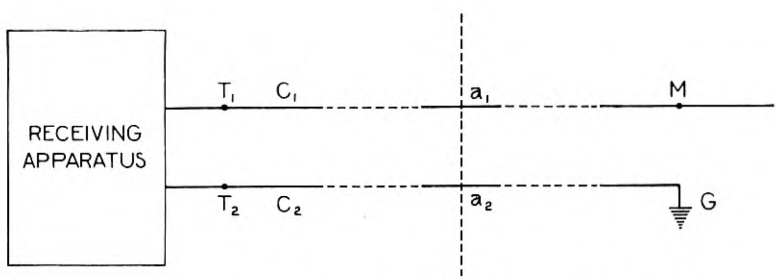
The method just described was applied to predetermine the interference at the terminals of the New York-Azores permalloy loaded cable. At the Azores terminal the cable reaches deep water within a few miles of the terminal, and the results indicated that the magnitude of interference to be expected would be sufficiently small to permit of signalling at the speed at which it was desired to operate. At the New York terminal, however, the ocean for a distance of about 100 nautical miles is comparatively shallow, and cables in this vicinity are exposed to rather severe disturbances. This is partly due to unusually strong

[7] See Milnor "Submarine Cable Telegraphy," Trans. *A. I. E. E.*, Vol. 41, p. 20, 1922.

stray fields from the numerous electric railway systems in the neighborhood of New York. By means of an amplifier and a recording string oscillograph records were obtained of the interference on the Western Union Telegraph Company's non-loaded cables terminating at New York. In taking these records a number of terminal networks were employed, with various attenuation characteristics, in order to obtain an idea of the distribution of interference with respect to frequency. Another series of tests was made, on board the Western Union cableship "Clowry," during which a cable was raised from deep water, cut, and interference studies made on the two parts of the cable. A study of these results according to the method that has just been described indicated that unless some means were employed for reducing the terminal interference, a great sacrifice of signal speed would have to be made, at least on westbound traffic. The remedy that was adopted is a special type of earth connection 100 miles at sea, to which the ground terminal of the receiving apparatus is connected. The theory of this arrangement will now be developed.

For the purpose of diminishing extraneous interference there is provided on most submarine cables an earthing arrangement, which, as shown diagrammatically in Fig. 5, consists of a core C_2 of the same general type as that used in the main cable C_1, and which may be

FIG 5



armored either with the main cable or in an independent sheath. This core usually extends for a distance of a few miles from the shore, to a point G, where the conductor of the core is grounded on the armor of the main cable. The receiving apparatus associated with the main cable conductor is then connected to earth through the sea earth conductor and the earth connection at its sea terminal. It is evident that if the main core and the sea earth core are close together they will both be exposed to the disturbances encountered between the terminal and the point where the sea earth conductor is grounded. If the two cores reacted in the same degree to these disturbances, then it is clear that corresponding to each disturbing impulse at T_1 due to pick-up at any point a_1 on T_1M there would be an equal impulse at T_2 due to pick-up at a_2 the corresponding point on $T_2\,G$ and there would be no resulting difference of potential impressed on the receiving network due to these disturbances. As a matter of fact the section $T_2\,G$ does not react to disturbances in the same manner as the section $T_1\,M$, even though the two cores have identical linear characteristics. Although the impedances looking landward from a_1 and a_2 will be equal, the impedances looking seaward from the two points are likely to be widely different, and the impedances into which electromotive forces induced at a_1 and a_2 work will not be equal. The same disturbance will therefore set up currents of different amplitudes in the two conductors, and there will be a difference of potential between T_1 and T_2 which will be indicated on the receiving instrument. Another way of looking at this effect is to consider the disturbances picked up at a_1 and a_2 as resulting in transient waves of potential and current which are propagated along the two conductors in both directions from the points of pick-up. The waves travelling from a_1 to T_1 are equal to the corresponding waves travelling from a_2 to T_2. A similar equality holds for the waves travelling from a_2 to G and from a_1 to M. On arriving at G the waves on the sea earth conductor are reflected and travel back along the conductor, finally arriving at T_2. Since there is no corresponding reflection on the main conductor, there will be an unbalanced disturbance, the magnitude of which depends upon the amount by which the disturbance was attenuated in travelling over the route a_2-G-T_2.

The remedy [8] for the condition just described is to eliminate reflection at the sea end of the sea earth conductor, or, if for any reason, there is a reflection at the point M, to balance it with an equal reflection at the point G. This can be done by grounding the sea earth conductor at G through a network having an impedance that bears the same relation to the impedance of the conductor $G\,T_2$ as the impedance of the cable seaward of M bears to the impedance of the conductor $M\,T_1$. When the two cores $T_1\,M$ and $T_2\,G$ are alike, the impedance of the network should equal the impedance of the main cable at M.[9]

[8] Osborne, U. S. Patent 1,390,580—1921.
Heurtley, Br. Patent 198,978—1923.
Gilbert, Br. Patent, 218,261—1926.

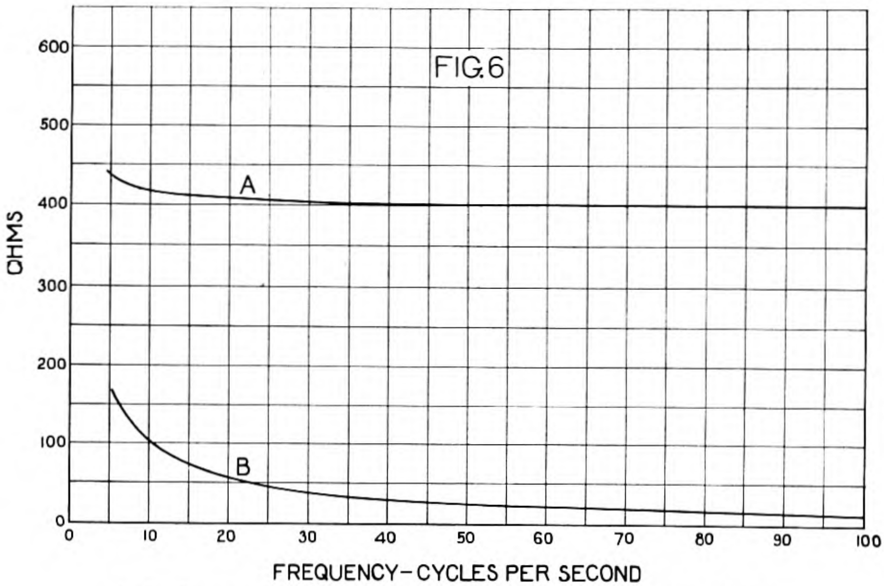
[9] There is one important type of disturbance which has not been dealt with in the preceding discussion, namely, that due to the signal currents on cables which cross or lie close to the cable in which we are interested. It is evident that the electromotive forces induced in the cable conductor due to such causes behave in the same manner as any other disturbing electromotive force and that the magnitude of their effect can be reduced by the use of a balanced sea earth conductor terminated at a point beyond the region of disturbance.

With this source of unbalance between the main core and the sea earth core removed or greatly reduced, it becomes increasingly important that the factors affecting the pick-up and the transmission of interference on the two cores be made as nearly as possible the same. In manufacturing the cable, core lengths should be paired off in such a manner that the electrical constants of any portion of the sea earth core match the constants of the corresponding portion of the main core, and the two cores should preferably be armored together.

By an extension of the method employed in deriving formula (5) an expression for the interference-susceptibility frequency characteristic of a cable having a balanced type of sea earth can be derived. This expression will consist of two terms, the first representing the resultant interference due to lack of balance between the sea earth conductor and the main core, and a second term, similar in form to (5), representing the interference picked up on the portion of the cable beyond the sea earth termination. Because of the difficulties involved in balancing, there is a value below which the first term cannot practically be reduced, which residue amounts to a few per cent of the magnitude of interference that would be encountered on this portion of the cable if the balanced type of sea earth were not employed. The second term can be reduced to any desired value by terminating the sea earth in water of sufficient depth. It is evident that when the sea earth has been extended to a point where the second term is small compared with the first, the limit of interference reduction is reached.

The question as to how far from shore the sea earth should be located in a particular case is an economic problem, the optimum location being that where the increase in value of the cable, due to diminution of interference by further extension of the sea earth, balances the additional cost of making the extension. In some cases it is found economical to obtain the desired ratio of signal-to-interference by means of a more efficient and expensive core rather than by an extended sea earth conductor. In the case of transatlantic cables terminated at points on the English Channel, or on the North Sea, for example, sea earth conductors several hundred miles in length are required in order to get a deep water termination. By increasing the weight of the main conductor, thereby increasing the amplitude of signals received over the cable, a greater amount of interference can be tolerated, in which case a comparatively short sea earth can be employed, just long enough to get rid of local interference and of the pick-up of signals from cables terminating nearby.

An inductively loaded submarine telegraph cable possesses characteristics which make the balanced type of sea earth particularly adaptable. Fig. 6 shows the real and imaginary parts of the characteristic impedance of a typical cable designed to operate at a speed corresponding to about 60 c.p.s. It is evident that for all frequencies above 20 c.p.s. the impedance can be approximated very closely by a pure resistance of about 400 ohms. In contrast to this, the character-



A. Resistance Component. *B*. Reactance Component (negative)

istic impedance of a non-loaded type of cable varies with frequency and has a reactance component about as large as the resistance component. In the case of the loaded cable the problem of designing a terminating network for the sea earth conductor is therefore comparatively simple, being a matter of finding a method of including in the cable structure a resistance of several hundred ohms. It is true that a network of this sort does not provide a good balance for frequencies much below 20 c.p.s., and components of interference of these low frequencies will be found at the cable terminals due to the lack of balance between the main cable and the sea earth. As was previously pointed out, however, these components will be so greatly attenuated by the signal correcting networks that their effect upon the receiving instrument will be inappreciable. This is illustrative of a general property of the loaded telegraph cable, namely, that when a cable is suitably designed for the frequency at which it is to be operated its characteristic impedance approximates closely to a resistance over a range of frequencies which extends considerably below the signal frequency, so that the resultant interference due to employing a resistance termination for the sea earth conductor will be attenuated to such a degree by the signal correction networks that it will in general have a negligible effect upon the receiving instrument. Moreover, it is probable that a considerable amount of low frequency disturbance is picked up beyond the sea earth and the gain obtained by improving the balance for these frequencies would not be very great.

A practical design for the terminating resistance consists of a length of several hundred feet of stranded wire, approximately 0.05 inch in diameter, of high resistivity material, insulated with gutta percha. After being joined at one end to the sea end of the sea earth core, the insulated conductor is served with jute and laid up with the main core for armoring exactly in the same manner as any other portion of the sea earth core. The free end of the conductor is grounded by connecting to the armor wires in the usual manner. A structure of this sort satisfies very completely the requirement of simplicity and lightness, and is as easily maintained as a length of ordinary cable similarly located.

There is a second characteristic of the loaded type of cable that tends to simplify the problem of the design of a balanced type of sea earth. It has been shown that the portion of the extraneous interference that it is most desirable to eliminate consists of the components of

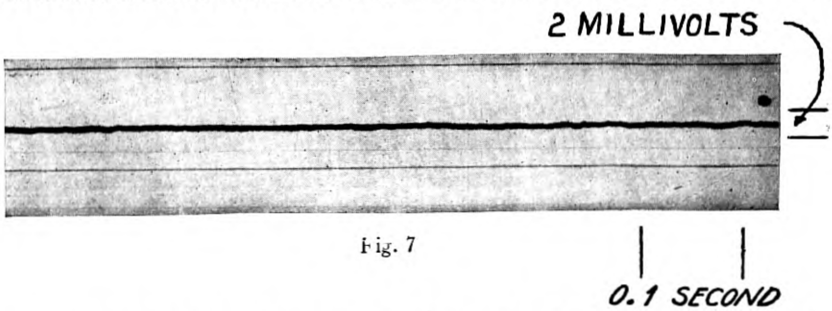


Fig. 7

frequencies in the neighborhood of the signal frequency. Since the operating speed of a loaded cable is five to ten times that of the corresponding non-loaded cable, it is evident from the preceding discussion that in order to effect a given reduction of interference in any particular locality, the sea earth of the loaded cable can be located closer to shore and in shallower water than in the case of the non-loaded cable.

In the case of the New York-Azores cable the balanced type of sea earth has been very effective in reducing extraneous interference. Fig. 7 is an oscillographic record of the terminal interference between this cable and its sea earth taken at the same time and under the same conditions as Fig. 1, which is the record of terminal interference on an adjacent non-loaded cable provided with the ordinary type of sea earth. In both cases a large condenser was inserted between the cable and the amplifier to reduce the "zero wander" due to components of very low frequency. Comparison of the two records indicates that the interference on the cable with the ordinary sea earth is about ten times that on the cable with the balanced sea earth. The contrast between the two types of sea earth is still more pronounced at times when terminal interference is unusually large. It has been found possible, for example, to operate the New York-Azores cable during violent local electrical storms when neighboring cables were compelled to cease operation.

Neutralization of Telegraph Crossfire

By R. B. SHANCK

Synopsis: With the simple means here described for neutralizing mutual interference between parallel telegraph circuits, it has been found practicable to effect a reduction to 10 or 20 per cent of the original values. This has improved considerably the operation of some circuits and made available others which were previously unsuitable. The resulting improvement in transmission has made possible the elimination of certain intermediate telegraph repeaters with material savings. The neutralizing apparatus has no material effect when crossfire is not present, that is, when the paralleling wires are idle. It has been found that the use of arrangements here described on certain long open wire circuits makes possible fast manual full-duplex operation where only medium-speed half-duplex operation was possible before. Furthermore, in the case of some cable circuits where it was impossible to operate more than two telegraph circuits per quad, it is now practicable to obtain four telegraph circuits.

Introductory

Many ground-return telegraph circuits are subject to serious mutual interference due to their proximity to one another on pole lines or in cable and in certain cases due to interconnection in office apparatus. The interfering currents, commonly referred to as "crossfire", in one telegraph circuit, caused by the transmission of signals on paralleling telegraph circuits, have caused considerable difficulty in the operation of such circuits. Crossfire has either limited the speed of operation or seriously impaired the quality of transmission in many cases.

In the following there are described methods which have been successfully applied to a number of ground-return polar-duplex telegraph circuits in the Bell System for the purpose of neutralizing crossfire. These arrangements are comparatively inexpensive and afford a marked improvement in transmission. This paper deals specifically with methods for use on wires which are either used simultaneously for telephone purposes or at least are grouped and transposed so as to be suitable for telephone operation; there is, however, no reason why the principles may not be profitably applied in many cases where wires are intended exclusively for telegraph use.

Nature of Crossfire

When mutual admittance, or coupling, exists between two telegraph circuits, operation of one, of course, occasions extraneous current impulses in the other circuit. The presence of such impulses in the receiving apparatus at the terminals of the disturbed circuit results in adverse effects on the telegraph signals. In the case of closely parallel circuits extending between two stations, considerable interference is generally experienced both at the station from which the disturbing signal is transmitted and also at the distant station. In this paper, the crossfire current (noted in the interfered—with circuit) at the station from which the interfering signal is sent will be referred to as "sending-end crossfire" and that at the distant station as "receiving-end crossfire". For example, assume two parallel wires from A to B; if a signal be sent on wire No. 1 from A to B, sending-end and receiving-end crossfire will appear in the receiving apparatus of wire No. 2 at A and B, respectively. This may mutilate incoming signals, or in extreme cases cause false signals.

The type of line circuit and the kind of apparatus employed have a considerable effect upon the amount of crossfire between circuits. It has been found that it depends chiefly upon the amount of mutual capacitance and, to a lesser extent, upon the natural mutual inductance of the wires; mutual conductance or leakage is responsible for some d-c. crossfire during periods of low insulation resistance but this increment is in general comparatively unimportant. As will be brought out later, loading[1] of circuits has a large effect on crossfire. Such factors as the gauge of wire, separation between wires, length of circuit and the presence of other wires on the same pole line have, of course, considerable influence.

In the Bell System plant, crossfire is in general of little consequence except among the four wires of a "phantom" group, the reasons for which will be discussed later. It is of interest to note that receiving-end crossfire is comparatively much more serious between wires in cable than between those of open-wire lines. Entrance cable, that is, cable employed to bring open-wire circuits into large cities, has comparatively little effect, as the length is generally short. Such apparatus as the composite sets which are used to derive d-c. telegraph circuits from telephone wires, and in certain cases filters used in connection with superposed carrier-current systems, contribute to crossfire inasmuch as they introduce some coupling, chiefly mutual capacitance.

Fig. 1 shows schematically the circuit arrangement of the polar-duplex telegraph apparatus in conjunction with a pair of wires composited for simultaneous telephone and telegraph operation. These types of apparatus are well known and will therefore be described only briefly. Independent two-way telegraph transmission is possible on each wire since the receiving relay occupies a position in a

[1] See "Development and Application of Loading for Telephone Circuits", Shaw and Fondiller, A. I. E. E. Jour. March, 1926.

balanced circuit analogous to that of a Wheatstone-bridge galvanometer as regards outgoing signals, and is therefore operated only by incoming signals. The fact that the home relay is not influenced by the home battery will be appreciated by considering the battery at the distant station to be short-circuited and the artificial line to balance the line and distant apparatus perfectly. When battery is introduced at the distant station however it causes a signal to be

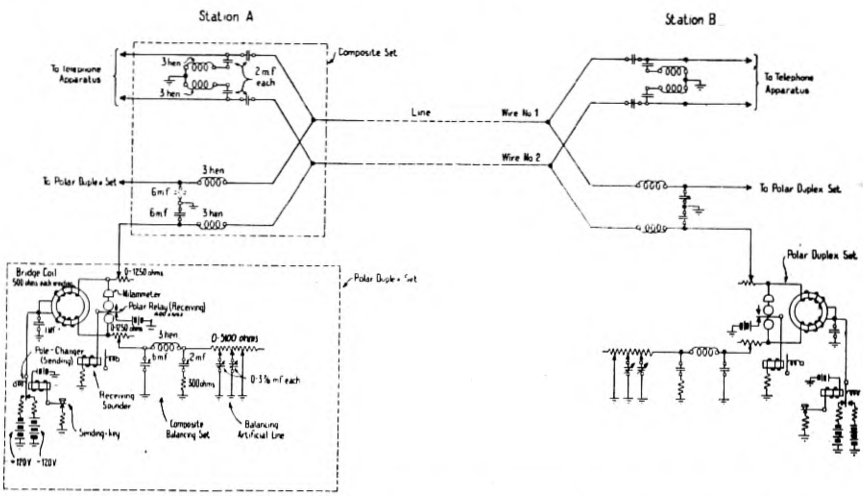


Fig. 1—Line equipped with composite sets and ground-return polar duplex sets

received at the home station as it corresponds to inserting a battery in one arm of a Wheatstone bridge. The "bridge coil" is connected so as to be series-aiding for incoming signals, the inductance being about 75 henries in this case, and parallel-opposed for outgoing signals, the inductance then being about 3 henries. The composite set serves to separate the telephone currents from those of the two telegraph circuits by "filtering" action or frequency discrimination, the telegraph employing a frequency range below that of the telephone.

The oscillograms which are shown in Fig. 2 illustrate the wave shape and magnitude of crossfire impulses in comparison with the normal operating currents, in a typical composited large gauge cable circuit. Trace A shows the wave-shape of the normal current in the line at the sending end (reduced to about one-seventh as compared to the other waves), and B shows the current in the receiving relay at the distant station. Although it is not outstanding in the

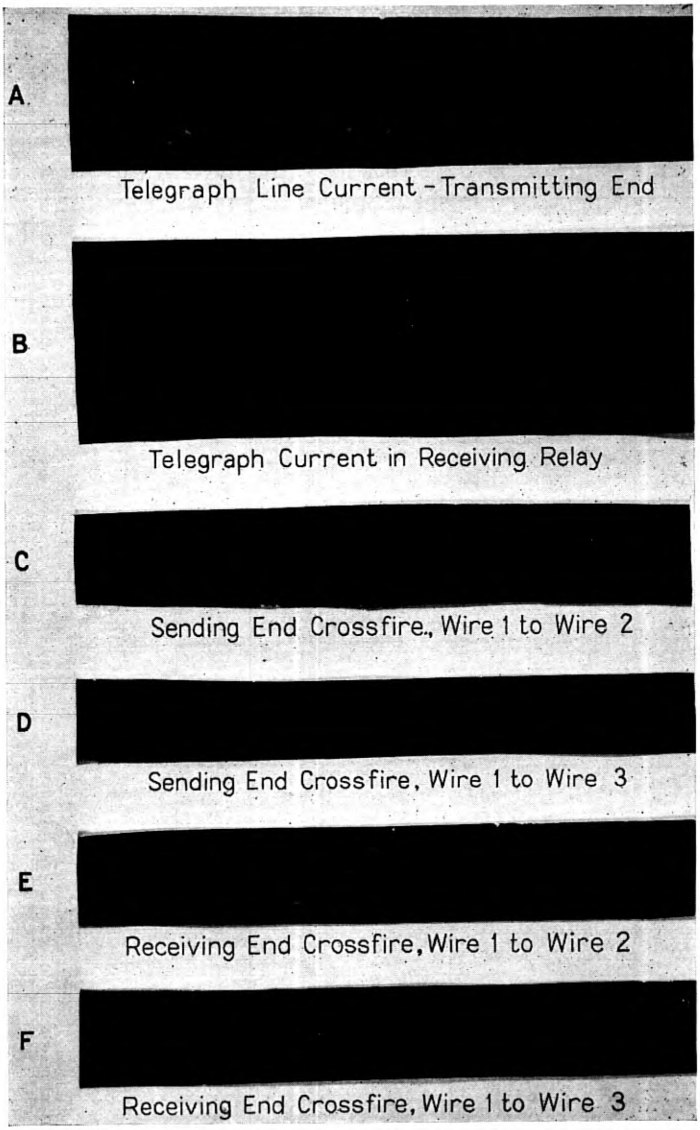


Fig. 2—Telegraph operating and crossfire currents. 13 B.&S. Ga. loaded cable Quad. 90 miles in length

Note 1: Oscillograms of crossfire current taken with vibrator in series with receiving relay

Note 2: Wave "A," 150 milliamperes per inch; other waves 20 milliamperes per inch present instance, the sending-end current is usually characterized by peaks and rapid changes, while the received wave is somewhat rounded off, and this results in most of the induction taking place in the portion of the line near the sending station and the apparatus at that station. C illustrates sending-end crossfire between wires of the same pair and D that between wires of different pairs but in the same quad. Trace E shows the receiving-end crossfire between wires of the same pair and F that between wires of different pairs but in the same quad. C, D, E, and F may be considered as superposed in various combinations on B to obtain an idea of the mutilation of signal waves at usual speeds of manual Morse operation.

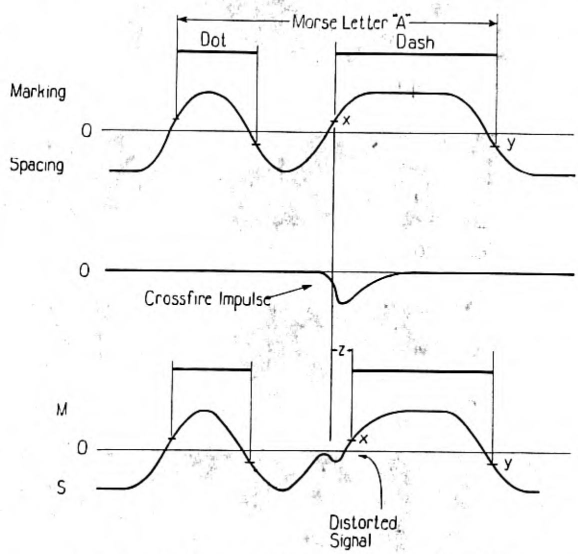


Fig. 3—Distorting effect of crossfire impulse

Fig. 3 has been drawn to illustrate how a crossfire impulse may cause distortion of a telegraph signal. The lowest wave is a combination of the received signal and crossfire impulse which are shown above. X and Y are the points at which the polar relay operates, assuming that it is required that the current build up or down appreciably beyond zero in order to move the armature. It will be clear that the dash has been shortened by the amount Z. Obviously only a limited amount of such distortion is allowable in telegraph signals. Under some conditions the crossfire is of sufficient strength to cause false signals, such as an extraneous dot in a long space or a break (space) in a dash.

An additional serious effect of crossfire is that it interferes with the obtaining of accurate duplex balance adjustment, since crossfire currents mask the effect of small changes in the balancing artificial line.

Principles of Neutralizing Arrangements

The principles involved in neutralizing the crossfire will first be discussed for the simplest case, that is, with only two parallel wires, reserving the case of four wires for the next section of this paper.

Sending-end Crossfire

An arrangement suitable for neutralizing the sending-end crossfire between two polar-duplex circuits is illustrated in Fig. 4. The

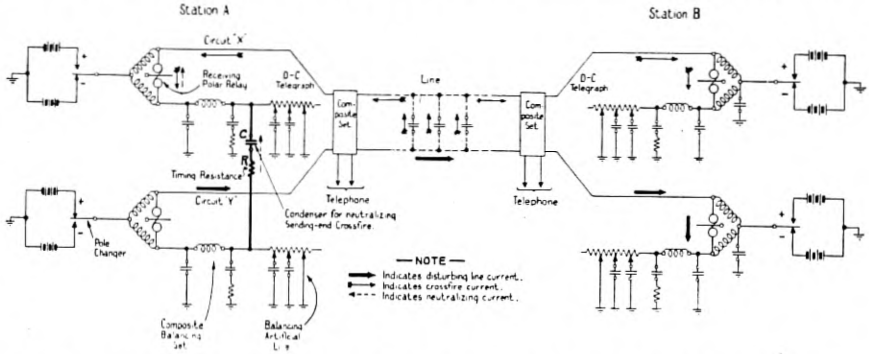


Fig. 4—Method of neutralizing sending-end crossfire between two telegraph circuits

heavy arrows indicate the disturbing line current which flows when the tongue of the pole-changer of circuit Y at station A moves from the negative to the positive pole. The feathered arrows show the direction of the resulting crossfire currents which tend to flow through the polar relays of circuit X. It will be apparent that the sudden increase, in a positive direction, of the potential applied to circuit Y would cause an impulse of current in the relay of X at the sending station A in the direction shown if the circuits were coupled by capacity only or by the natural mutual inductance of the two parallel ground-return circuits. Neutralization is effected by providing a mutual admittance between the two balancing artificial lines to simulate that existing between the real lines. It will be clear that upon the operation of the pole-changer of circuit Y, an impulse will pass through the neutralizing circuit C, R, and through the relay of circuit X at A in such direction as to oppose the crossfire current.

(The neutralizing impulse is indicated by the dotted arrows.) Another point of view is that a symmetrical or balanced arrangement similar to a Wheatstone bridge is provided in which the coupling of the line circuits is balanced by the coupling introduced between the artificial lines. It has been found experimentally that a simple connection consisting of a condenser and a timing resistance in series as shown are sufficient to effect neutralization on either open-wire or cable circuits. It will, of course, be seen that such a connection is effective for neutralizing crossfire from either circuit into the other, and furthermore that it is capable of performing both functions simultaneously.

As shown in Fig. 4 the neutralizing connection is made at the beginning of the artificial line (at the junction of it and the composite balancing set). This is a convenient point and has been found satisfactory for the purpose.

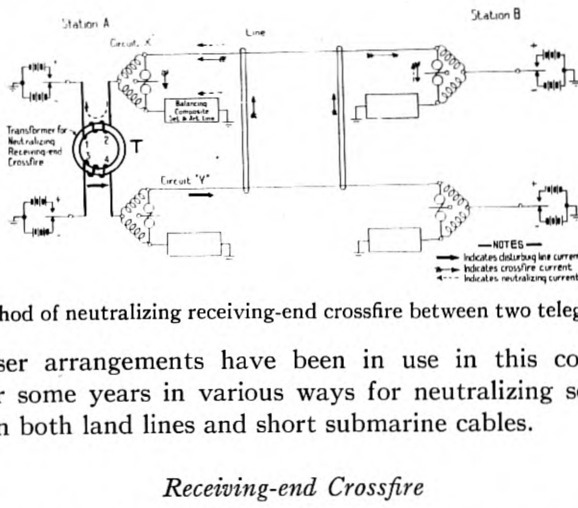


Fig. 5—Method of neutralizing receiving-end crossfire between two telegraph circuits

Condenser arrangements have been in use in this country and abroad for some years in various ways for neutralizing sending-end crossfire on both land lines and short submarine cables.

Receiving-end Crossfire

For neutralizing receiving-end crossfire use is made of special connections at the sending end. The method consists in impressing a neutralizing impulse on the disturbed circuit at the sending station, in such manner as not to affect incoming signals at that station, (that is, it does not introduce sending-end crossfire); the neutralizing impulse will then travel along the interfered-with circuit so as to arrive at the distant station at the time when the crossfire impulse appears at that station.

The operation of the receiving-end crossfire neutralizing apparatus will be made clear by reference to Fig. 5, in which the heavy arrows indicate the disturbing current, the feathered arrows the crossfire current and the dotted arrows the neutralizing current. The last mentioned current is impressed upon the disturbed circuit X by means of a transformer connection (T) between the "apex" or transmitter branches of the two circuits. It will be obvious that if a good duplex balance has been obtained the neutralizing impulse will divide practically equally between the real and artificial lines of circuit X and substantially none of it will pass through the receiving polar relay of circuit X, on account of the balanced bridge arrangement. It will therefore have no effect on signals received at

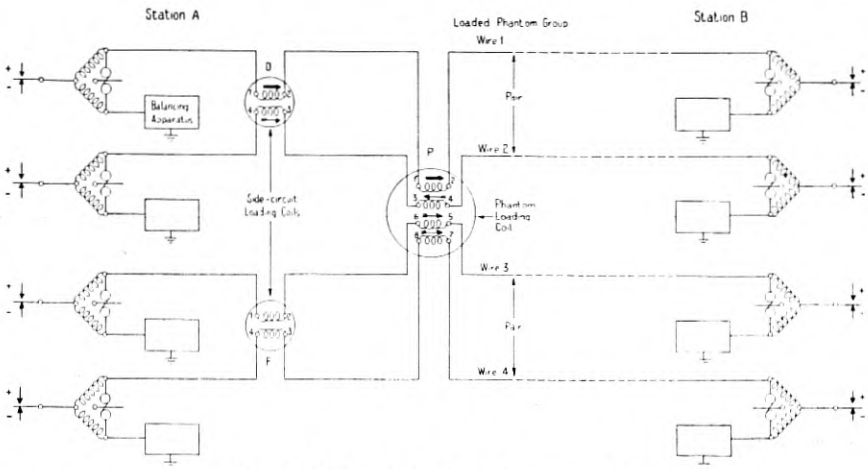


Fig. 6—Effect of loading coils on crossfire

Note: → Signalling Current ➜ Induced Current

A, but will generate neutralizing impulses which will travel over X to B so as to appear at B at the same time as the crossfire currents. It has been found possible to employ coupling such that the receiving-end crossfire is practically eliminated in the polar relay. The proper poling of the neutralizing transformer has been found to be as indicated in Fig. 5 for all types of circuit to which the device has been applied. The crossfire impulse has the direction shown, for the reason that capacity coupling predominates.

Line Characteristics

The first part of this section will be devoted to a discussion of the effect of loading and line transpositions. This will show why, in the telephone plant, it is necessary to deal with crossfire among the four wires of a phantom group only. Then arrangements for use with a group of four wires employing the principles explained above in connection with the case of two parallel wires, will be covered.

It is of interest to consider the effect of the loading coils which are employed in conjunction with many telephone lines.[2] In Fig. 6, coils D and F represent side-circuit loading coils on pairs 1–2 and 3–4, respectively, and P a phantom-circuit loading coil. Such coils are connected into telephone circuits at intervals to introduce inductance into the two telephone side circuits and the phantom telephone circuit, respectively.

The action of the side-circuit coil, (D), will first be considered. If a positive telegraph impulse is sent from A to B over wire 1, as indicated by the heavy arrow, it is evident that the coil acting as a transformer will set up a crossfire current in wire 2 in the same geographical direction, as indicated by the feathered arrow. The relation of this impulse to those due to capacity coupling is of interest since the capacity effect predominates. Comparison with Fig. 5, will show that at the transmitting station the impulse due to coil D will oppose the sending-end crossfire which is due to capacity coupling between circuits, while at the distant end it will augment the receiving-end crossfire due to capacity coupling. Coil F functions similarly in pair 3–4.

In the case of the phantom loading coil (P) sending an impulse from A to B on wire 1 results in disturbing currents in the same geographical direction in wires 3 and 4 and in the opposite direction in wire 2, since the coil is connected so that two windings are series-opposed in each side circuit and parallel-aiding in the phantom circuit. Comparing with Fig. 5, as before, it will be seen that for wires of a group but not of the same pair these coils tend to neutralize the sending-end crossfire which is due to mutual capacitance and augment the receiving-end crossfire due to capacity coupling; the conditions will be reversed however for wires of the same pair.

In the case of loaded circuits, crossfire, therefore, is due to loading as well as to the mutual capacitance and inductance of the wires and the coupling which exists in office apparatus, so that the final result is difficult to predict. Work with loaded circuits, which has been largely confined to cables indicates that on such circuits receiving-end crossfire is generally greater than sending-end crossfire, and sending-end crossfire between wires which are in the same phantom group but not in the same pair is so small as to be almost negligible.

Line transposition of telephone circuits has been discussed at

[2] Shaw and Fondiller, loc. cit.

considerable length in a previous paper.[3] Such transpositions consist in interchanging systematically the pin positions of the two wires of a pair and of the wires of the two pairs comprising a phantom circuit. It should be clearly understood that while these transpositions are effective in balancing a two-wire or metallic circuit against other circuits, they cannot be used to balance ground-return circuits (such as the telegraph circuits in question) against each other; however, their effect in varying the separation of the different wires from each other has a great influence on the coupling between the ground-return circuits.

A possible transposition section for an open-wire phantom group is shown in Fig. 7. It will be seen that the two wires of a pair are

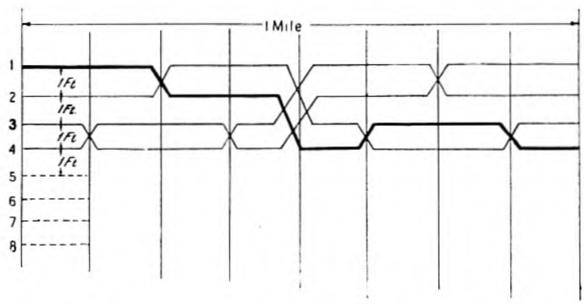


Fig. 7—Line transpositions of open-wire phantom group

always adjacent to each other and will therefore have considerable coupling; a wire of one pair is adjacent to a particular wire of the other pair for only one-fourth of the distance, and wires of the two pairs will therefore have much less coupling.

A brief consideration will make it clear that coupling between wires of separate phantom groups is comparatively small. Each wire of the group 1 to 4 occupies pin position 4 only one-fourth of the distance, and if 5 to 8 be phantomed each wire of the latter group will use pin position 5 one-fourth of the distance. It follows that a wire of group 1 to 4 will be adjacent to a particular wire of group 5 to 8 only one-sixteenth of the distance in a long circuit. If 5-6 be non-phantomed however each wire of the pair will use position 5 half of the time and will be adjacent to each wire of 1 to 4 one-eighth of the way. The next crossarms above and below are each two feet distant and carry wires transposed so as to minimize the coupling.

[3] "The Design of Transpositions for Parallel Power and Telephone Circuits," H. S. Osborne, Proc. A. I. E. E. 1918, Vol. XXXVII p. 739.

It should be noted that in addition to the reduction in coupling due to increasing the spacing there is a large reduction due to shielding when a third conductor is interposed between two others.

In the case of cable circuits the wires are twisted in groups of four so as to be transposed practically continuously. On account of the

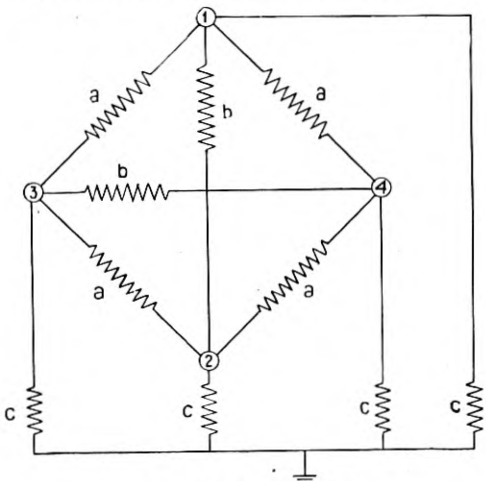


Fig. 8—Admittance Network

smaller separation, mutual capacitances and the resulting crossfire among wires of a phantom group, are considerably greater than in open wire.

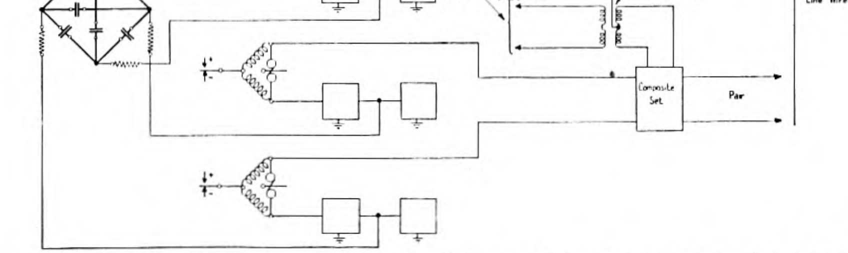


Fig. 9—Condenser arrangement for neutralizing sending-end crossfire between telegraph circuits on a phantom group

The result of transposing is that for practical purposes in connection with the crossfire problem the other wires of the line can be ignored and a phantom group represented by a network of admittances as shown in Fig. 8, where 1 and 2 represent a pair and 3 and 4 the other pair.

A network of the form of Fig. 8, is used as shown in Fig. 9, for neutralizing sending-end crossfire among the four wires of a phantom

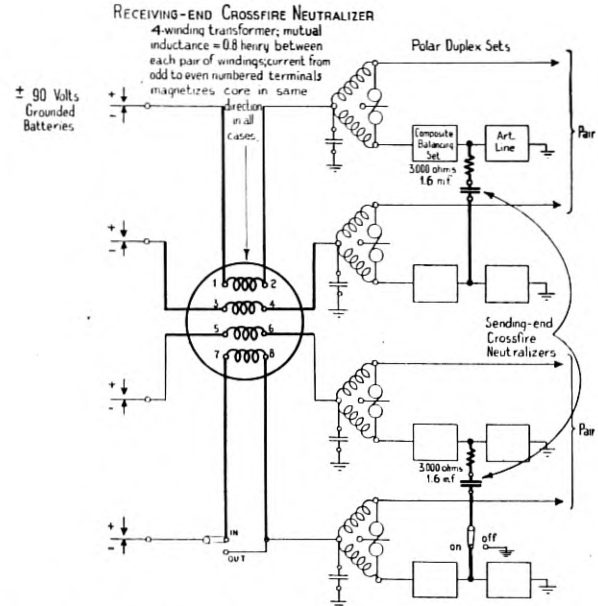


Fig. 10—Arrangement for neutralizing crossfire between telegraph circuits on a loaded No. 13 B.&S. Ga. phantom group 90 to 120 miles long in cable

group. The grounded branches shown in Fig. 8 are omitted, however, since the duplex artificial lines themselves constitute these branches. The six-mesh network consists of condensers, the timing resistances when used being external to the network as a matter of convenience.

For coupling together the apex circuits of the four wires to neutralize receiving-end crossfire, it is possible to use a special four-winding transformer analogous to the six-condenser network, but where an ordinary transformer as shown in Fig. 10 will not suffice it is convenient to employ two or three transformers in combination. For example a two-winding transformer may be added to each pair of the arrangement illustrated in Fig. 10 so as to provide additional coupling between the two wires of a pair.

It will readily be seen that with neutralizers applied at each end of the four circuits, transmission of signals on one of them will generate the proper impulses for neutralizing both sending and receiving-end crossfire from that circuit into the other three. Furthermore, neutralization will take place with all wires operating simultaneously in either or both directions.

Application to Different Circuits

It is in general not practicable to compute the constants of the neutralizing devices, but this is unnecessary since it is an easy matter to determine them experimentally. In making trials to determine the proper amount of capacity and inductance required to neutralize crossfire effectively, it is fortunately possible to design the various parts independently of each other to a considerable extent. For example, the diagonals of the six-condenser network may be determined after the condensers in the sides have been approximated very roughly, or vice-versa; likewise, the amount of inductance required between each pair of circuits may be approximated independently, but if a single coil is to be used for coupling more than two circuits, all the circuits should be connected up in making the test. Sending and receiving-end crossfire may, of course, be treated separately.

It is convenient to vary the capacity of the condensers, but not usually the inductance of the transformer. In the latter case a coil with excess mutual inductance may be used together with a variable resistance shunt.

In order to design neutralizing arrangements or to determine whether or not they are effective, tests may readily be made by observing the deflection of a milliammeter connected in series with the polar relay of a bridge polar-duplex set while signals are sent on the parallel circuit. In a similar way a differential meter may be used in a differential duplex set. A somewhat more accurate test may be made by observing the response of the receiving relay, preferably with variable electrical bias. The disturbing signals are of course sent from the same station in checking sending-end crossfire and from the distant station in checking receiving-end crossfire.

Representative anti-crossfire capacity values are given in the following table for No. 8 B.W.G. (0.165 in., 2.5 mm.) composited open-wire copper circuits, 300 to 500 miles (500 to 800 km.) in length and No. 12 A.W.G. (0.104 in., 1.5 mm.) circuits 150 to 300 miles (250-550 km.) in length. No timing resistance is required usually. In practice there are material variations from one circuit to another.

Gauge	Loading	Non-Phantomed Pair	Phantom Group Network[4] Diagonals	Sides
8	Non-loaded	1.7 mf.	1.7 mf.	1.2 mf.
8	Loaded	1.1 "	1.1 "	0.55 "
12	Non-loaded	1.1 "	1.1 "	0.8 "
12	Loaded	0.8 "	0.8 "	0.4 "

The superposition of carrier-current channels by means of filters connected on the drop side of the d-c. composite set of course has no appreciable effect on crossfire. However, the use of "transfer filters" at intermediate points to transfer the carrier from one pair to another increases the coupling between wires of a pair and this may be taken care of by increasing the capacity of the diagonals of the condenser network.

The arrangement shown in Fig. 10 has been found to be suitable for use with 90 to 120 mile (145 to 190 km.) sections of No. 13 B.&S. gauge (0.072 in., 1.8 mm.) loaded cable circuits.

In the case of open-wire circuits, receiving-end crossfire is commonly not serious excepting in special cases where high-frequency carrier telephone or telegraph transfer filters are employed. In such cases, effective neutralization may be secured by coupling the wires of each pair by means of a transformer, no such coupling being provided between the wires of separate pairs of a phantom group.

In some cases the neutralizing arrangements and the telegraph repeaters have been wired to jacks in such a manner that it is possible to patch the neutralizers from set to set by means of cords when the line assignment is changed temporarily. In some cases it is not desirable to provide such elaborate arrangements and, therefore, switches are provided for disconnecting the neutralizing apparatus from each set independently. In the case of the condensers, the duplex balance of the other telegraph sets associated with the particular group of condensers is preserved by switching directly to ground the connection from the artificial line of the set to be disconnected, as shown in conjunction with the lowermost duplex set in Fig. 10. Switches for disconnecting the neutralizing transformers are illustrated also for the same duplex set in this figure.

Practical Results Obtained with Neutralization

The following table gives data which show roughly the amount of crossfire between wires of a phantom group without neutralizing

[4] See Figure 9.

arrangements for various circuit conditions. It will be noted that crossfire between a pair of wires used for a telephone side circuit is considerably greater than that between wires of a phantom group but not of the same pair. This is in accord with what was brought out above regarding coupling. The receiving-end crossfire is much greater between cable circuits than in the case of open wires, due to the greater mutual capacitance and heavier loading.

Crossfire Current in Per Cent. of Operating Direct-Current

For Average Repeater Sections

Type of Circuit	Sending End: From Other Wire of Pair	Sending End: From Wire Of Other Pair	Receiving End: From Other Wire of Pair	Receiving End: From Wire Of Other Pair
Non-loaded Open Wire..............	20	10	10	5
Loaded Open Wire.	10	5	5	5
13 B.&S.Ga. Loaded Cable............	20	5	30	25

It is practicable to reduce the crossfire to 10 or 20 per cent. of the original value by means of the arrangements which have been described. This has improved considerably the operation of some circuits and made available others which were unsuitable for use. By improving transmission so as to avoid the use of intermediate telegraph repeaters material savings have been effected in certain cases.

The neutralizing apparatus has no material effect on the quality of telegraph transmission obtained when crossfire is not present, that is, with the parallel wires idle; the application of them, however, reduces greatly the detrimental effect of crossfire on transmission. For example, the use of these arrangements on certain long open-wire circuits makes possible fast manual full-duplex (two-way) operation where only medium-speed half-duplex (one-way) operation was possible before. Furthermore, in the case of some cable circuits where it was previously impossible to operate more than two telegraph circuits per group of four wires, it is now practicable to obtain four telegraph circuits per quad.

Due to reduction of crossfire, it is usually possible to secure a much better duplex balance after the neutralizers have been applied. The application of anti-crossfire condensers however requires that a somewhat different setting of the duplex artificial line be obtained for the best balance, since the extra connection has appreciable admittance to ground.

Operation of Thermionic Vacuum Tube Circuits

By F. B. LLEWELLYN

SYNOPSIS: Given the static characteristic of grid current-grid potential, and plate current-plate potential, for any three element vacuum tube, the general exact equations for the output current when the tube is connected in circuits of any impedance whatsoever, and excited by any variable voltage, are here derived. The method of derivation is illustrated in the special case where resistances only are considered, and the adaptation of complex impedance to use in non-linear equations is shown. Approximations that are allowable in various practical applications are indicated, and the equations are applied in some detail to grid-leak detectors, and in brief to other types of detectors, modulators, amplifiers and oscillators.

Certain repetitions of previous work are contained in these pages, as it is believed that the applications of the novel features introduced are illustrated thereby better than by a description dealing only with new material.

THE equations in use at the present time for the relation between input voltage and output current in thermionic vacuum tubes are those developed by a number of pioneers in Radio Communication. They have been summarized very concisely, and somewhat extended in an important paper by John R. Carson, entitled "A Theoretical Study of the Three Element Vacuum Tube," which appeared in the Proceedings of the Institute of Radio Engineers, April, 1919. For some time past the need of relations which include the effect of the variation of certain quantities, considered constant in Mr. Carson's paper, has been growing. Especially in the case of detection and modulation has this need become pronounced. Moreover, in the special case of grid leak detectors, the need for a general theoretical analysis has not, to the author's knowledge, been completely satisfied.

PURPOSE

It is, therefore, the purpose of the present paper to derive general exact equations for the output current from a three-element thermionic vacuum tube when it is connected in circuits of general impedance, both on the input and output sides, and to show specific methods of applying these general equations to several special cases, with emphasis on the case of the grid leak detector. It is also proposed to show that, whether used for detectors, modulators, amplifiers, or oscillators, the same fundamental theory applies. It is hoped that the theory and methods given will form a basis upon which a complete rational design of vacuum tube circuits may be built.

THEORY

In the derivation of these equations, no limitations whatever should be imposed. Consider a three-element vacuum tube connected in circuits of general impedance on both input and output sides. The grid is allowed to take convection current. The amplification factor, μ, is considered variable, and the effect of plate potential on grid current is included. Under these conditions, the total plate current of the tube can merely be said to be a function of the grid and plate potentials; and the total grid current, likewise, is some other function of the grid and plate potentials. The fundamental relations:

$$I_p = I_p(E_g, E_p) \tag{1}$$

$$I_g = I_g(E_g, E_p) \tag{2}$$

express, the operation of the device. They represent the static characteristics of the tube. It is from these two relations alone that the general theory must be built.

In order to do this, the following notation will be employed:

$$\left.\begin{aligned} I_p &= I_{po} + i_p \\ I_g &= I_{go} + i_g \\ E_p &= E_{po} + e_p \\ E_g &= E_{go} + e_g \end{aligned}\right\} \tag{3}$$

It will be recognized that the lower case letters represent variations in the normal values of the currents and voltages denoted by the zero subscripts. It should be noted, moreover, that all voltages and currents refer to the effect directly on the element of the tube, plate or grid as the case may be.

With the aid of (3), equations (1) and (2) may be written

$$i_p = P_1 e_g + P_2 e_p + \tfrac{1}{2} P_3 e_g^2 + P_4 e_g e_p + \tfrac{1}{2} P_5 e_p^2 + \ldots \tag{4}$$

$$i_g = T_1 e_g + T_2 e_p + \tfrac{1}{2} T_3 e_g^2 + T_4 e_g e_p + \tfrac{1}{2} T_5 e_p^2 + \ldots \tag{5}$$

where the P's and T's have the following significance:

$$\left.\begin{aligned} &P_1 = \frac{\partial I_{po}}{\partial E_g} \quad P_2 = \frac{\partial I_{po}}{\partial E_p} \quad P_3 = \frac{\partial^2 I_{po}}{\partial E_g^2} \quad P_4 = \frac{\partial^2 I_{po}}{\partial E_g \partial E_p} \quad P_5 = \frac{\partial^2 I_{po}}{\partial E_p^2} \\ &T_1 = \frac{\partial I_{go}}{\partial E_g} \quad T_2 = \frac{\partial I_{go}}{\partial E_p} \quad T_3 = \frac{\partial^2 I_{go}}{\partial E_g^2} \quad T_4 = \frac{\partial^2 I_{go}}{\partial E_g \partial E_p} \quad T_5 = \frac{\partial^2 I_{go}}{\partial E_p^2} \end{aligned}\right\} \tag{6}$$

Equations (4) and (5) are obtained directly from the extension of Taylor's Theorem. The P's may be written in more useful form with the aid of (1) and the well-known definitions of the amplification factor, μ, the plate resistance, r_p, and the grid resistance, r_g. Thus, from (1)

$$\left.\begin{aligned} &\mu=\left.\frac{\frac{\partial I_p}{\partial E_g}}{\frac{\partial I_p}{\partial E_p}}=-\frac{dE_p}{dE_g}\right]_{I_p}\\ &\frac{1}{r_p}=\frac{\partial I_p}{\partial E_p}\\ &\frac{1}{r_g}=\frac{\partial I_g}{\partial E_g}\end{aligned}\right\}\text{(by definition)} \tag{7}$$

Hence

$$\left.\begin{aligned} &P_1=\frac{\mu}{r_p}\\ &P_2=\frac{1}{r_p}\\ &P_3=\frac{1}{r_p}\frac{\partial\mu}{\partial E_g}+\frac{\mu}{r_p}\frac{\partial\mu}{\partial E_p}-\mu^2\frac{r_p'}{r_p^2}\\ &P_4=\frac{1}{r_p}\frac{\partial\mu}{\partial E_p}-\mu\frac{r_p'}{r_p^2}\\ &P_5=-\frac{r_p'}{r_p^2}\end{aligned}\right\} \tag{8}$$

where

$$r_p'=\frac{\partial r_p}{\partial E_p}.$$

In similarly treating the T's, it was found convenient to introduce an entirely new symbol. This has been done with reluctance, for it is realized that considerable difficulty has been experienced in the standardization of symbols already in use. But inasmuch as the simplification of both physical interpretation and mathematical expression which results from the use of this new symbol is enormous, its addition is believed to be warranted.

This new symbol we will call the reflex factor, and will denote it by the symbol, ν. It is analogous in its effect on the grid circuit to the effect of μ on the plate circuit. Its definition is analogous to that of μ. Thus, from (2):

$$\nu=\left.\frac{\frac{\partial I_g}{\partial E_g}}{\frac{\partial I_g}{\partial E_p}}=-\frac{dE_p}{dE_g}\right]_{I_g} \tag{9}$$

Comparison of (7) and (9) shows that while μ is equal to minus the ratio of the increments of E_p and E_g necessary to maintain the plate current constant, ν is equal to minus the ratio of the increments of E_p and E_g necessary to maintain the grid current constant. On the other hand, while in the case of μ, the ratio

$$\left.\frac{dE_p}{dE_g}\right]_{I_p}$$

is intrinsically negative and occurs in (7) with a negative sign, making μ intrinsically a positive number; in the case of ν, the ratio

$$\left.\frac{dE_p}{dE_g}\right]_{I_g}$$

is usually intrinsically positive, and occurs in (9) with a negative sign; hence ν is usually intrinsically a negative number.

With the foregoing definition, the T's may be written as follows:

$$\begin{aligned}
T_1 &= \frac{1}{r_g} \\
T_2 &= \frac{1}{\nu r_g} \\
T_3 &= -\frac{r_g'}{r_g^2} \\
T_4 &= \frac{1}{r_g}\frac{\partial}{\partial E_g}\left(\frac{1}{\nu}\right) - \frac{1}{\nu}\frac{r_g'}{r_g^2} \\
T_5 &= \frac{1}{\nu r_g}\frac{1}{\partial E_g}\left(\frac{1}{\nu}\right) + \frac{1}{r_g}\frac{\partial}{\partial E_p}\left(\frac{1}{\nu}\right) - \frac{1}{\nu^2}\frac{r_g'}{r_g^2}
\end{aligned} \tag{10}$$

where $$r_g' = \frac{\partial r_g}{\partial E_g}.$$

The effective value of T_2, when taken over a cycle of sine wave form, has sometimes been called the reflex mutual conductance (L. A. Hazeltine), and has been denoted by g_n. An attempt to adapt this notation to the present purpose has not proved feasible. For reference, it may be noted in the limiting case, where the amplitude of the sine wave approaches zero:

$$g_n = \frac{1}{\nu r_g}.$$

With the relations given thus far, the problem may now be more specifically stated as follows:

It is desired to express i_p, the output current through a general impedance in the plate circuit, as an explicit function of e, a variable voltage applied in series with a general impedance in the grid circuit.

Special Case

The following special case will make the detailed derivation, where complex quantities are considered, more intelligible.

For this special case consider a vacuum tube connected in circuits containing only pure resistances. Let the resistance in the grid circuit be denoted by Q and that in the plate circuit be denoted by Z. Fig. 1 illustrates this circuit. Let i_p and i_g be determined to satisfy the following series:

$$i_p = a_1 e_g + a_2 e_g^2 + \dots \tag{11}$$

$$i_g = b_1 e + b_2 e^2 + \dots \tag{12}$$

(11) and (12) are valid since (4) and (5) are formally power series. As seen from Fig. 1, e represents a variable voltage impressed in series with the resistance, Q, on the grid of the tube.

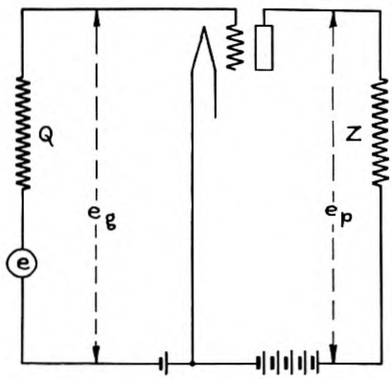


Fig. 1—Fundamental circuit diagram

These equations will give the plate and grid currents as explicit function of the voltages e_g and e, respectively, if we can evaluate the a's and b's. To do this, we have the relation

$$e_p = -i_p Z. \tag{13}$$

Substituting (11) in (4) and equating coefficients of like powers of e_g, we may evaluate a_1 and a_2 and thus express i_p as an explicit function of e_g:

$$i_p = \frac{\mu}{(r_p+Z)} e_g + \frac{1}{2}\left[\frac{-\mu^2 r_p r_p' + \mu \dfrac{\partial \mu}{\partial E_p}(r_p^2 - Z^2) + \dfrac{\partial \mu}{\partial E_g}(r_p+Z)^2}{(r_p+Z)^3}\right] e_g^2. \tag{14}$$

In equation (14), when the amplification factor, μ, is considered constant, we have the well-known relation as given in Mr. Carson's paper

$$i_p = \frac{\mu e_g}{(r_p+Z)} - \frac{1}{2}\frac{\mu^2 r_p r_p'}{(r_p+Z)^3} e_g^2 + \cdots \tag{15}$$

Experiments have shown, however, that when the resistance, Z, is not small compared to r_p the modulation resulting from variations of μ amounts to an appreciable part of the total. When the grid is maintained at a negative potential with respect to the filament, (14) may be simplified somewhat by the relation which then holds quite closely[1], namely:

$$\mu \frac{\partial \mu}{\partial E_p} = \frac{\partial \mu}{\partial E_g}.$$

Equation (14) then becomes

$$i_p = \frac{\mu}{r_p+Z} e_g - \frac{1}{2}\left[\frac{\mu^2 r_p r_p'}{(r_p+Z)^3} - \frac{2r_p \frac{\partial \mu}{\partial E_g}}{(r_p+Z)^2}\right] e_g^2 + \cdots \tag{16}$$

This equation is applicable to the calculation of the output current when e_g is known, and the grid takes no convection current, as is the case in very many circuits met with in practice.

It is instructive to investigate the relative magnitudes of the two components of the second term of (16) in an actual experimental case. For convenience, the contribution of the second component of this term will be called, "μ modulation." A vacuum tube was measured and found to have the following properties under operating conditions

$$r_p = 6400$$
$$r'_p = -61.3$$
$$\mu = 5.84$$
$$\frac{\partial \mu}{\partial E_g} = .05$$

The results of applying these to (16) are shown in the following table:

Z	Total modulation	μ modulation, %
0	$.03341/10^3\ e^2$	23.35
r_p	$.00515/10^3\ e^2$	37.8
$2\ r_p$	$.00186/10^3\ e^2$	46.6
$4\ r_p$	$.000889/10^3\ e^2$	55.0

[1] See appendix I for proof of this.

This illustrates strikingly the importance of the variation of μ in modulators and detectors.

Equation (14) is expressed in terms of e_g, the voltage directly on the grid of the tube. We may derive the expression for i_p in terms of e, a voltage impressed in series with a resistance, Q, in the external grid circuit by noting that

$$e_g = e - i_g Q.$$

Hence, from (12),

$$e_g = (1 - b_1 Q)e - b_2 Q e^2 + \dots \tag{17}$$

Therefore

$$i_p = a_1(1 - b_1 Q)e - [a_1 b_2 Q - a_2(1 - b_1 Q)^2]e^2 + \dots \tag{18}$$

and, as in (13),
$$e_p = -i_p Z.$$

Substituting (17) and (18) into (5) and equating coefficients of like powers of e, we get

$$b_1 = \frac{T_1 - T_2 a_1 Z}{1 + T_1 Q - T_2 a_1 Z Q}, \tag{19}$$

$$b_2 = \frac{[-a_2 Z T_2 + \frac{1}{2} T_3 - a_1 Z T_4 + \frac{1}{2} a_1^2 Z^2 T_5](1 - b_1 Q)^2}{1 + T_1 Q - T_2 a_1 Z Q}. \tag{20}$$

The T's may be expressed in terms of r_g and ν with the aid of (10). The complete solution of this special case for first and second order effects is then given by (18) above, in which we have now evaluated the a's and b's.

Mathematical Digression

Before the detailed steps in the complete development of the general case, with general impedances instead of resistances, are attempted, the following digression on the use of complex quantities in non-linear equations is apposite. Included at this point, it serves a two-fold purpose; first, the notation to be used is illustrated by means of simple applications; second, it calls to mind the fundamental ideas involved in the representation of impedances by complex quantities.

Consider a current, I. If periodic, this current may be represented by a Fourier series and expressed as the sum of a number of cosine terms. Thus

$$I = I_h\left(\frac{\epsilon^{j(ht+\phi)} + \epsilon^{-j(ht+\phi)}}{2}\right) + I_k\left(\frac{\epsilon^{j(kt+\theta)} + \epsilon^{-j(kt+\theta)}}{2}\right) + \dots \tag{21}$$

where the symbol, j, represents the imaginary, $\sqrt{-1}$. For brevity this may be written

$$I = (i_{1h} + \bar{i}_{1h}) + (i_{1k} + \bar{i}_{1k}) + \dots \tag{22}$$

where the bar over a symbol denotes the conjugate imaginary of the same symbol unbarred. If this current flows through a circuit containing resistance, self-inductance, and capacity in series, we have

$$e=RI+L\frac{dI}{dt}+\frac{1}{C}\int Idt. \tag{23}$$

Substituting for I its equivalent, as given by (21) or (22), we may write the result in abbreviated form as follows:

$$e=(z_h i_{1h}+\bar{z}_h\bar{i}_{1h})+(z_k i_{1k}+\bar{z}_k\bar{i}_{1k})+\text{ - - -} \tag{24}$$

where

$$z_n=R+Ljn+\frac{1}{Cjn},$$

$$\bar{z}_n=R-Ljn-\frac{1}{Cjn}.$$

When the current flows through a network of impedances, we may always write the equivalent series impedance of the network. Hence equation (24) may be extended to cover the general case. It will be noted that lower case z's have been used to represent impedances in the above discussion. Throughout this paper the attempt has been made to employ the lower case letters to denote quantities which involve time, reserving the capitals for those which do not involve time. With this understanding, Z denotes a resistance, while z represents a general impedance, which, of course, varies with the time variation of a voltage impressed on it. With the aid of (24) we are in a position to treat non-linear equations by the complex method. Thus, omitting conjugates e^2 becomes,

$$\begin{aligned}e^2=z_h{}^2 i_{2(2h)}+z_k{}^2 i_{2(2k)}+2z_h\bar{z}_h i_{2(0h)}+2z_h z_k i_{2(h+k)}\\ +2z_h\bar{z}_k i_{2(h-k)}+2z_k\bar{z}_k i_{2(0k)}+\text{ - - -}\end{aligned} \tag{25}$$

which may be written

$$e^2=e_{2(2h)}+e_{2(2k)}+e_{2(0h)}+e_{2(h+k)}+e_{2(h-k)}+e_{2(0k)}+\text{ - - -} \tag{26}$$

In (25) and (26) the significance of the double subscript notation is brought out. The first symbol in the subscript refers to the order of the term, and the second refers to the frequency.

In the light of the foregoing discussion, the problem of writing the general equations for the thermionic vacuum tube may be attacked.

General Analysis

Coming back to the detailed problem in hand, we follow out the method illustrated in the special case, but must use the notation developed in the preceding section to take care of a *general* impedance, z, in the plate circuit, and a *general* impedance, q, on the grid circuit. Fig. 1 as before, shows the skeleton circuit, where, however, lower case z and q must be substituted for the capitals. Then

$$e = e_{1h} + \bar{e}_{1h} + e_{1k} + \bar{e}_{1k} + \cdots + e_{1n} + \bar{e}_{1n}. \tag{27}$$

Analogous to (11) and (12):

$$\left.\begin{aligned} i_p = {} & a_{1h} e_{g1h} + \bar{a}_{1h} \bar{e}_{g1h} + a_{1k} e_{g1k} + \bar{a}_{1k} \bar{e}_{g1k} + \cdots \\ & + a_{2(h-k)} e_{g2(h-k)} + a_{2(h-k)} e_{g2(h-k)} + \cdots \end{aligned}\right\} \tag{28}$$

$$i_g = b_{1h} e_{1h} + \bar{b}_{1h} \bar{e}_{1h} + b_{1k} e_{1k} + \bar{b}_{1k} \bar{e}_{1k} + \cdots$$

$$+ b_{2(h-k)} e_{2(h-k)} + \bar{b}_{2(h-k)} \bar{e}_{2(h-k)} + \cdots$$

Hence, analogous to (13) and (17):

$$e_p = -\Sigma[a_{1n} z_{1n} e_{g1n} + \bar{a}_{1n} \bar{z}_{1n} \bar{e}_{g1n} + a_{2m} z_m e_{g2m} + \bar{a}_{2m} \bar{z}_m \bar{e}_{g2m}] \tag{29}$$

$$e_g = \Sigma[(1 - b_{1n} q_n) e_{1n} + (1 - \bar{b}_{1n} \bar{q}_n) \bar{e}_{1n} - b_{2m} q_m e_{2m} - \bar{b}_{2m} \bar{q}_m \bar{e}_{2m}] \tag{30}$$

where the summation refers to terms of different frequencies but of similar form.

From this point on, the procedure is exactly the same as that given in the special case. Coefficients of terms of like order *and frequency* are equated, and the final results are:

$$\left.\begin{aligned} i_p = {} & \Sigma a_{1h}(1 - b_{1h} q_h) e_{1h} \\ & + \Sigma[(1 - b_{1h} q_h)^2 a_{2(2h)} - a_{1(2h)} q_{(2h)} b_{2(2h)}] e_{2(2h)} \\ & + \Sigma[(1 - b_{1h} q_h)(1 - b_{1k} q_k) a_{2(h+k)} - a_{1(h+k)} q_{(h+k)} b_{2(h+k)}] e_{2(h+k)} \\ & + \Sigma[(1 - b_{1h} q_h)(1 - \bar{b}_{1k} \bar{q}_k) a_{2(h-k)} - a_{1(h-k)} q_{(h-k)} b_{2(h-k)}] e_{2(h-k)} \\ & + \Sigma[(1 - b_{1h} q_h)(1 - \bar{b}_{1h} \bar{q}_h) a_{2(0h)} - a_{1(0h)} q_{(0h)} b_{2(0h)}] e_{2(0h)} \\ & + \ldots \end{aligned}\right\} \tag{31}$$

where the summation refers to terms of different frequencies but of similar form. Note that, having $a_{2(h-k)}$ and $b_{2(h-k)}$, we may readily write the appropriate expressions for the other a_2's and b_2's by reference to the formation in equation (31).

In (31) the a's and b's are given by:

$$\left.\begin{aligned}
a_{1h} &= \frac{\mu}{r_p+z_h} \\
a_{2(h-k)} &= \frac{\frac{1}{2}\left[-\mu^2 r_p r_p' + \mu\frac{\partial\mu}{\partial E_p}(r_p^2 - z_h\bar{z}_k) + \frac{\partial\mu}{\partial E_g}(r_p+z_h)(r_p+\bar{z}_k)\right]}{(r_p+z_h)(r_p+\bar{z}_k)(r_p+z_{h-k})} \\
b_{1h} &= \frac{1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}}{r_g+q_h\left(1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}\right)} \\
b_{2(h-k)} &= \frac{\left\{\begin{aligned}&\frac{1}{2}\left[-r_g r_g'\left(1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}\right)\left(1-\frac{\mu}{\nu}\frac{\bar{z}_h}{r_p+\bar{z}_k}\right)-2a_{2(h-k)}\frac{r_g^2}{\nu}z_{2(h-k)}\right. \\ &-\frac{\partial}{\partial E_g}\left(\frac{1}{\nu}\right)\left(\frac{\mu z_h r_g^2}{r_p+z_h}+\frac{\mu\bar{z}_k r_g^2}{r_p+\bar{z}_k}-\frac{\mu^2 r_g^2}{\nu}\frac{z_h\bar{z}_k}{(r_p+z_h)(r_p+\bar{z}_k)}\right) \\ &\left.+\frac{\partial}{\partial E_p}\left(\frac{1}{\nu}\right)\left(\frac{r_g^2\mu^2 z_h\bar{z}_k}{(r_p+z_h)(r_p+\bar{z}_k)}\right)\right]\end{aligned}\right\}}{\left[r_g+q_h\left(1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}\right)\right]\left[r_g+\bar{q}_k\left(1-\frac{\mu}{\nu}\frac{\bar{z}_k}{r_p+\bar{z}_k}\right)\right]\left[r_g+q_{(h-k)}\left(1-\frac{\mu}{\nu}\frac{z_{(h-k)}}{r_p+z_{(h-k)}}\right)\right]}
\end{aligned}\right\}\quad(32)$$

Discussion of General Equations

Equations (31) and (32) contain the general solution of the problem. The formulas are too long to consider all effects at one time but if we separate (31) into components and consider each component separately, useful applications may be secured.

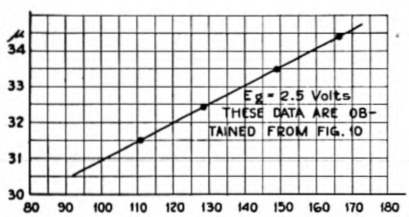
First taking the component that gives rise to amplification effects, we get

$$\begin{aligned} i_{p(h)} &= a_{1h}(1-b_{1h}q_h)e_{1h} \\ &= \left(\frac{\mu}{r_p+z_h}\right)\left(\frac{r_g}{r_g+q_h\left(1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}\right)}\right)e_{1h}. \end{aligned}\qquad(33)$$

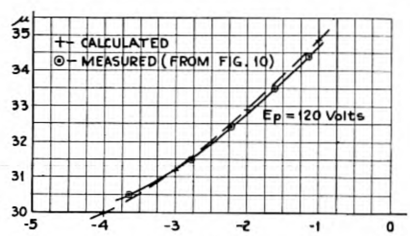
The point to be noted in this relation is that when $q_h << r_g$ we have the well-known relation

$$i_{p(h)} = \frac{\mu}{r_p+z_h}e_{1h}.\qquad(34)$$

Since amplifiers are usually operated under the condition that r_g is exceedingly large, the general solution has contributed nothing new to the amplifier equations for conditions where the grid is maintained at a negative potential with respect to the filament. But for positive values of grid potential both q_h and the reflex factor, ν, enter into the calculations. It may be remarked in passing that when the grid and plate are both positive by the same amount, the absolute value of ν is approximately equal to, or somewhat less than, μ. On the other hand, when, as is usually the case, the plate potential is much greater than the positive grid potential, the magnitude of ν is much greater than μ.



E_p, Volts



E_g, Volts

Fig. 2—Change of μ with plate and grid potentials. The points on the calculated curve were obtained as follows: since

$$\frac{\partial \mu}{\partial E_g} = \mu \frac{\partial \mu}{\partial E_p}$$

then

$$\mu = \frac{\mu_o + KE_p}{1 - KE_g}$$

where

$$K = \frac{\partial \mu}{\partial E_p} \div \left(1 + E_g \frac{\partial \mu}{\partial E_p}\right), \ \mu_o = \frac{\mu - E_p \dfrac{\partial \mu}{\partial E_p}}{1 + E_g \dfrac{\partial \mu}{\partial E_p}}.$$

From the upper curve, for $E_p = 120$, $E_g = -2.5$

$$\mu = 32, \ \frac{\partial \mu}{\partial E_p} = .05328,$$

whence

$$K = .06146, \mu_o = 29.55$$

We next consider the component of (31) that results in plate curvature detection or modulation. It is given by

$$i_{p2}=(1-b_{1h}q_h)\,(1-\bar{b}_{1k}\bar{q}_k)a_{2(h-k)}\,e_{2(h-k)}$$

$$=\frac{\dfrac{r_g^2}{2}\left[-\mu^2 r_p r_p'+\mu\dfrac{\partial\mu}{\partial E_p}(r_p^2-z_h\bar{z}_k)+\dfrac{\partial\mu}{\partial E_g}(r_p+z_h)(r_p+\bar{z}_k)\right]e_{2(h-k)}}{\left[r_g+q_h\left(1-\dfrac{\mu}{\nu}\dfrac{z_h}{r_p+z_h}\right)\right]\left[r_g+\bar{q}_k\left(1-\dfrac{\mu}{\nu}\dfrac{\bar{z}_k}{r_p+\bar{z}_k}\right)\right](r_p+z_h)(r_p+\bar{z}_k)(r_p+z_{(h-k)})} \tag{35}$$

For rough calculations, μ may be regarded as a constant. For very careful work, this assumption should never be made without first drawing the curves of $\mu-E_g$ and $\mu-E_p$ and verifying the validity of the assumption under operating conditions. Examples of such curves are given in Fig. 2. When μ may be regarded as constant, and when the grid is maintained negative with respect to the filament, (35) becomes

$$i_{p(h-k)}=\frac{-\frac{1}{2}\mu^2 r_p r_p' e_{2(h-k)}}{(r_p+z_h)(r_p+\bar{z}_k)(r_p+z_{h-k})}$$

which may be put into the form given in Mr. Carson's paper, referred to before.

The third and last component of (31) is that which produces grid detection or modulation; namely

$$i_{p(h-k)}=-a_{1(h-k)}q_{(h-k)}b_{2(h-k)}e_{2(h-k)}$$

$$=\left(\frac{\mu q_{(h-k)}}{r_p+z_{(h-k)}}\right)\frac{1}{2}\left[-r_g r_g'\left(1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}\right)\left(1-\frac{\mu}{\nu}\frac{\bar{z}_k}{r_p+\bar{z}_k}\right)\right.$$

$$-\frac{2a_{2(h-k)}r_g^2 z_{(h-k)}}{\nu}+\frac{\partial}{\partial E_p}\left(\frac{1}{\nu}\right)\left(\frac{r_g^2\mu^2 z_h\bar{z}_k}{(r_p+z_h)(r_p+\bar{z}_k)}\right)-\frac{\partial}{\partial E_g}\left(\frac{1}{\nu}\right)$$

$$\left.\left(\frac{\mu z_h r_g^2}{r_p+z_h}+\frac{\mu\bar{z}_k r_g^2}{r_p+\bar{z}_k}-\frac{\mu^2 r_g^2 z_h\bar{z}_k}{\nu(r_p+z_h)(r_p+\bar{z}_k)}\right)\right]$$

$$\div\left[r_g+q_h\left(1-\frac{\mu}{\nu}\frac{z_h}{r_p+z_h}\right)\right]\left[r_g+\bar{q}_k\left(1-\frac{\mu}{\nu}\frac{\bar{z}_k}{r_p+\bar{z}_k}\right)\right]$$

$$\left[r_g+q_{(h-k)}\left(1-\frac{\mu}{\nu}\frac{z_{(h-k)}}{r_p+z_{h-k}}\right)\right] \tag{36}$$

In using this relation, ν may nearly always be considered constant. As grid leak detectors are often used, q consists of a resistance, R_g, and a condenser, C, in parallel. The values of R_g and C are so adjusted that the impedance of the combination to the first order frequencies is practically that of the condenser alone and may be neglected, and to the desired second order, or detected, frequency it is practically that of the resistance alone. When this is the case, and when the impedance in the plate circuit is a pure resistance, R_p, we may write (36) as follows:

$$i_{p3}=\frac{1}{2}\frac{\mu R_g\left(r_g r_g'+\frac{2a_{2m}r_g^2R_p}{\nu}\right)}{(r_p+R_p)r_g^2(r_g+R_g)}e_{2m}$$

and, considering μ constant, in order to obtain a physical view of the result, we get

$$i_{p3}=\frac{\frac{1}{2}\mu R_g\left[r_g r_g'-\frac{\mu^2 r_p r_p'}{(r_p+R_p)^3}\left(\frac{r_g^2R_p}{\nu}\right)\right]e_{2m}}{(r_p+R_p)r_g^2(r_g+R_g)} \tag{37}$$

This equation shows a condition that is present in many grid-leak detectors and which, it is thought, has not been generally appreciated. The condition referred to is the presence of the term involving the curvature of the plate characteristic in the grid detection component. This effect is *in addition* to the plate detection effect, given by (35). The plate detection component and the grid detection component are opposite in phase. Hence, it would seem that for best operation as a grid-leak detector, the curvature of the plate characteristic should be zero. This means a rather large value of E_b, the plate battery potential. In practice, however, it is usual to operate with fairly low values of E_b. The second term of the numerator of (37) accounts for this. It will be seen that detection resulting from this term and from the first term are in phase, since ν is intrinsically negative. Hence, it is entirely possible in certain cases for the optimum operating point to be such that the effect of the plate curvature is appreciable.

We now combine once more the three components, (33), (35) and (36), under the simplifying assumptions that μ and ν are constant and that ν is large enough so that terms containing ν in the denominator may be neglected. The result is

$$\left.\begin{aligned} i_p=&\frac{r_g}{(r_g+q_h)}\frac{e_{1h}}{(r_p+z_h)}+\frac{r_q}{(r_q+q_k)}\frac{e_{1k}}{(r_p+z_k)}+\ \dots\\ &+\left\{\frac{r_g^2}{(r_g+q_k)(r_g+\bar{q}_k)}\left[\frac{-\frac{1}{2}\mu^2 r_p r_p'}{(r_p+z_h)(r_p+\bar{z}_k)(r_p+z_{h-k})}\right]\right.\\ &\left.+\frac{\mu q_{(h-k)}}{(r_p+z_{h-k})}\left[\frac{\frac{1}{2}r_g r_g'}{(r_g+q_h)(r_g+\bar{q}_k)(r_g+q_{h-k})}\right]\right\}e_{2(h-k)}+\ \dots \end{aligned}\right\} \tag{38}$$

The first two terms of (38) are the amplification terms and represent undistorted reproduction in the plate circuit of the voltage, e, applied in the grid circuit. The third term of (38) represents the second order effects resulting from the curvature of the characteristics of the vacuum tube. The first part of this term represents the effects of so-called plate curvature detection or modulation, and the second part represents the effects of detection and modulation in the grid circuit. It is with this last-named component that the present paper is most concerned.

The Grid-Leak Detector

Fig. 3 shows the usual circuit diagram for a grid-leak detector. It is evident that the impedance, q, in this example is composed of the parallel combination of R_g and C. Suppose that the "h" and "k" frequencies are both radio frequencies, and that, for them, the

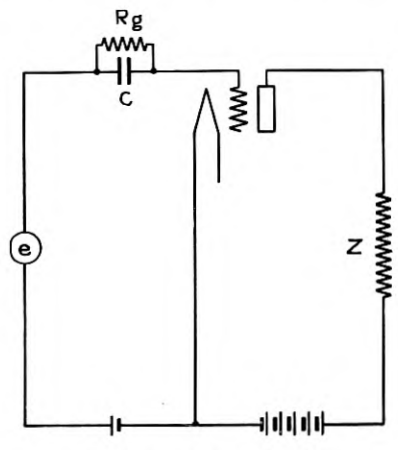


Fig. 3—Grid-leak detector

impedance offered by the resistance and condenser combination is practically that of the condenser, alone. Suppose, further, that practically the only impedance offered by the external circuit to the "(h-k)" frequency is that of the resistance, R_g, alone. This, of course, assumes that the "h-k" frequency is quite low. Then, when E_b, the voltage of the plate battery, is such that r_p' is very small, and when

$$e = A \cos ht + B \cos kt$$

we have, from (38), for second order effects

$$
\left.\begin{aligned}
i_p=\tfrac{1}{2}r_g r_g'\left(\frac{\mu}{r_p+R_p}\right)\Bigg[&\frac{R_g}{\left(r_g+\frac{1}{jhC}\right)\left(r_g-\frac{1}{jhC}\right)(r_g+R_g)}\frac{A^2}{2}\\
&+\frac{\frac{1}{2jhC}}{\left(r_g+\frac{1}{jhC}\right)^2\left(r_g+\frac{1}{2jhC}\right)}\frac{A^2}{2}\cos 2ht\\
&+\frac{R_g}{\left(r_g+\frac{1}{jkC}\right)\left(r_g-\frac{1}{jkC}\right)(r_g+R_g)}\frac{B^2}{2}\\
&+\frac{\frac{1}{2jkC}}{\left(r_g+\frac{1}{jkC}\right)^2\left(r_g+\frac{1}{2jkC}\right)}\frac{B^2}{2}\cos 2kt\\
&+\frac{\frac{1}{j(h+k)C}}{\left(r_g+\frac{1}{jhC}\right)\left(r_g+\frac{1}{jkC}\right)\left(r_g+\frac{1}{j(h+k)C}\right)}AB\cos (h+k)t\\
&+\frac{R_g}{\left(r_g+\frac{1}{jhC}\right)\left(r_g-\frac{1}{jkC}\right)(r_g+R_g)}AB\cos (h-k)t\Bigg]
\end{aligned}\right\}\qquad (39)
$$

While most of the frequencies in this expression are unimportant in relation to any practical case on hand, they are included here to show the complete result for a given simple case. The last term of the above expression results in what is known as detection.

Let us consider this component in more detail as regards detection of an incoming modulated radio wave of the form

$$e=A\ (1+B\cos qt)\cos pt. \qquad (40)$$

They may be written

$$e=A\cos pt+\frac{AB}{2}\cos (p+q)t+\frac{AB}{2}\cos (p-q)t. \qquad (41)$$

If we identify the "p" frequency with "h," and let "k" have the values $(p+q)$ and $(p-q)$ in turn, the detection term of (39) gives

$$
\begin{aligned}
i_d=\tfrac{1}{2}r_g r_g'\left(\frac{\mu}{r_p+R_p}\right)\Bigg[&\frac{R_g}{\left(r_g-\frac{1}{jpC}\right)\left(r_g+\frac{1}{j(p+q)C}\right)(r_g+R_g)}\\
&+\frac{R_g}{\left(r_g+\frac{1}{jpC}\right)\left(r_g-\frac{1}{j(p-q)C}\right)(r_g+R_g)}\Bigg]\frac{A^2B}{2}\cos qt. \qquad (42)
\end{aligned}
$$

Reference to the mathematical digression will make clear the formation of the impedances in this expression. (42) is an important relation since it shows that there is a possibility that the amplitude of the detected current may be affected by the phase displacements of the side bands of the original wave which occur during the detection. For an ideal grid-leak detector, the magnitudes of the quantities $\frac{1}{pC}$, $\frac{1}{(p+q)C}$ and $\frac{1}{(p-q)C}$ are very small compared with r_g. Equation (42) then becomes

$$i_d = \frac{r_g'}{r_g} \frac{\mu}{(r_p+R_p)} \frac{R_g}{(r_g+R_g)} \frac{A^2B}{2} \cos qt. \tag{43}$$

In (43) we have the simplest possible form of the equation for a grid leak detector. The next step is to show methods for evaluating the quantities r_g and r_g'. As may be seen from the relations given in (7) and (8)

$$\frac{1}{r_g} = \frac{\partial I_{go}}{\partial E_g}, \qquad r_g' = \frac{\partial r_g}{\partial E_g},$$

and, since the action of the grid-leak detector depends upon r_g', it is evident that r_g is not a constant but varies with the value of E_g. We may obtain r_g by direct dynamical measurements or by drawing tangents to the static grid-potential grid-current curve of the tube under consideration. The value of r_g thus obtained applies only to a given value of E_g. Now E_g is a function of the voltage, e, as will be shown:

When e has the form given in (41), one of the resulting currents in the plate circuit is a direct current given by

$$i_{pd} = \frac{1}{2} \frac{r_g'}{r_g} \frac{\mu R_g}{(r_g+R_g)(r_p+R_p)} \left(\frac{A^2}{2} + \frac{A^2B^2}{2}\right).$$

This means that a constant voltage given by

$$e_{gd} = \frac{1}{2} \frac{r_g'}{r_g} \frac{R_g}{(r_g+R_g)} \left[\frac{A^2}{2} + \frac{A^2B^2}{2}\right] \tag{44}$$

must have appeared on the grid in order to produce the constant component of the plate current. This constant voltage is in addition to that which we have denoted by E_{go}, since it is part of e_g. Moreover, its intrinsic value is usually negative, since r_g' is usually negative. This means that the "effective" E_{go} has been reduced by the amount given in (44). However, r_g is slightly different at this new value of E_{go} and hence e_{gd} is not quite what a first calculation would lead one to believe. The method of arriving at the correct value for e_{gd}, and hence for r_g and r_g' is one of trial and error, for, after several recalculations of e_{gd} have been made, it will be found that check results are secured. Then r_g and r_g' may be determined from this resulting value of E_{go}.

In actually making these measurements, a dynamical method of measuring r_g will usually be found superior to the method of drawing tangents to the static characteristic, for the grid-potential grid-current characteristic of any tube is rather elusive because of the

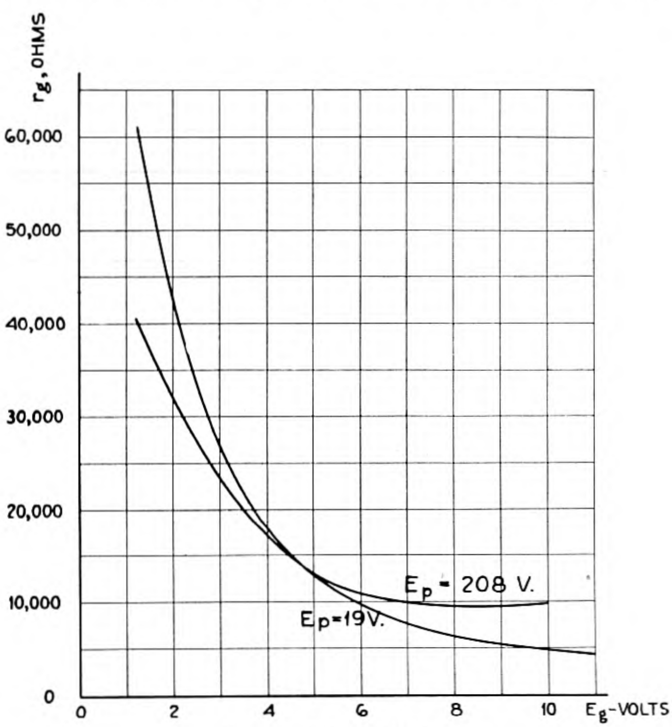


Fig. 4—Grid resistance

very small values of current involved. In the dynamic method a Wheatstone bridge circuit excited by a high frequency buzzer will be found convenient. The value of r_g' is, of course, obtained by drawing tangents to the r_g curve. Several examples of $r_g - E_g$ curves are shown in Fig. 4.

It must be recognized that, for large values of buzzer excitation, the dynamic value of r_g differs somewhat from that found by drawing tangents to the static characteristic. The dynamic value more nearly approaches the value r_g would assume with large signal inputs than does the static value. Hence, if a large signal input, e, is to be used, the amplitude of the buzzer excitation voltage on the grid should equal this amplitude as nearly as possible.

When the method of drawing tangents to the static characteristic is employed, a very close approximation to the value of r_g to use for large signal amplitudes may be obtained by drawing, not true tangents but secant lines to the static characteristic, which join points on the characteristic corresponding to the extreme, or peak, values of e_g.

When either method is used to obtain r_g, the value of r_g' must be obtained by drawing tangents to an $E_g - r_g$ curve.

With the precautions just given, and when the assumptions made in equation (43) are justifiable, an accuracy within 10% is easily obtained. While this is not very exact, nevertheless, it is a real advance over calculations made without taking the precautions just discussed for measuring r_g.

In many vacuum tubes the value of r_g is so high that the input impedance of the tube, resulting from the interelectrode capacities of the elements cannot justifiably be neglected. In order to include this effect, the following relations are applicable.

Consider the circuits shown in Fig. 5. This gives the equivalent circuit diagram for a vacuum tube with general impedances, z_1 and

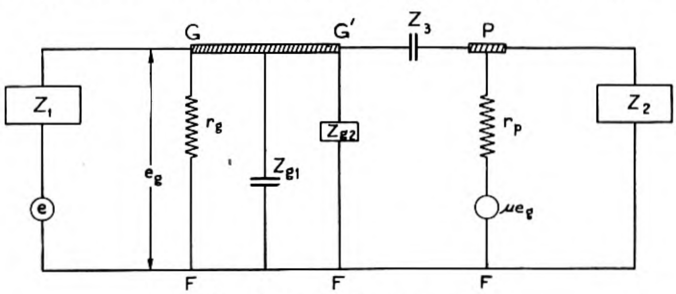


Fig. 5—Equivalent network

z_2, attached to the grid and plate, respectively. The plate to filament capacity may conveniently be included in z_2. The impedance, z_{g_2}, is the effective impedance of the network looking to the right from the point $G'F$. z_{g_1} is the grid to filament capacity of the tube, and z_3 is the grid to plate capacity.

We may write

$$z_g = \frac{z_{g_1} z_{g_2}}{z_{g_1} + z_{g_2}}.$$

In order to apply the general equations we must evaluate z_n and q_n.

To Evaluate z_n.

From the general equations, we have

$$i_p = \frac{\mu e_g}{r_p + z_n}.$$

Hence we may write Kirchoff's law for the plate circuit. This gives

$$i_p = \frac{e_g[(\mu+1)z_2 + \mu z_3]}{r_p z_2 + z_3(r_p + z_2)}.$$

Upon equating the two expressions for i_p, there results

$$z_n = \frac{z_2(\mu z_3 - r_p)}{(\mu+1)z_2 + \mu z_3}.$$

To Evaluate q_n.

By the general equations, we have

$$i_g = \frac{e}{r_g + q_n x}$$

where x stands for

$$\left(1 - \frac{\mu}{\nu}\frac{z_n}{r_p + z_n}\right).$$

This may be written

$$i_g = \frac{\frac{e}{x}}{\frac{r_g}{x} + q_n}$$

which says that Kirchoff's law may be applied to the grid circuit provided we use a modified voltage, $\frac{e}{x}$, and a modified grid resistance, $\frac{r_g}{x}$. Hence

$$i_g = \frac{\frac{e}{x} z_g}{(z_1 + z_g)\left(z_g + \frac{r_g}{x}\right) - z_g^2}.$$

Upon equating the two expressions for i_g there results

$$q_n = \frac{z_1}{z_g}\left(\frac{r_g}{1 - \frac{\mu}{\nu}\frac{z_n}{r_p + z_n}} + z_g\right).$$

To sum up; the following relations are applicable when interelectrode capacities or other coupling impedances are to be included:

$$z_g = \frac{z_{g1} z_{g2}}{z_{g1} + z_{g2}} \tag{45}$$

$$z_{g1} = \frac{1}{j\,\omega\, C_{gf}} \tag{46}$$

$$z_{g2} = \frac{z_2 z_3 + r_p(z_2 + z_3)}{z_2(\mu + 1) + r_p} \tag{47}$$

$$z_n = \frac{z_2(\mu z_3 - r_p)}{(\mu + 1) z_2 + \mu z_3} \tag{48}$$

$$q_n = \frac{z_1}{z_g} \left(\frac{r_g}{1 - \frac{\mu}{\nu} \frac{z_n}{r_p + z_n}} + z_g \right). \tag{49}$$

With the aid of (45), (46), (47), (48), (49), equation (42) may be modified to include all cases where the plate current resulting from detection or modulation in the grid circuit is desired, provided an accuracy greater than about 10% is not required. Where greater accuracy is essential, curves must be made to give the effect of the small terms in the numerator of the expression for b_{2m} in equation (36).

Before leaving the subject of grid-leak detectors, we will discuss briefly one of the physical aspects of grid-leak detection that the example just given, and the equations on which it is based, have emphasized. This is the fact that the fiction of the time-constant of the grid-leak and condenser combination is not a necessary physical interpretation of the phenomena which occur in the grid circuit. Indeed, in many cases, the time constant method of calculating the leak and condenser gives quite erroneous and misleading results. These cases occur when the impedance looking into the vacuum tube is of such value, as it often is, that the magnitudes and forms of q_{1n} and q_{2m} are materially changed from those which they would have if z_g were neglected, and when r_g is not large compared with q_n and q_m. Equation (38) shows that, for greatest plate current resulting from grid detection, q_h and q_k should be as small as possible, while $q_{(h-k)}$ should be as large as possible. It is, then, a filter problem, and if treated as such, will give reliable results both as to physical interpretation and numerical values.

In the special case when the input and detected frequencies are $\frac{h}{2\pi}$ and $\frac{s}{2\pi}$, respectively, and where $z_g \gg r_g$:

$$q_h = \frac{1}{jhC}$$

$$q_s = \frac{\frac{R}{jsC}}{R + \frac{1}{jsC}}$$

R = leak resistance

C = capacity in parallel with R

Then, the optimum size for the condenser, C, is easily shown to be

$$C^2 = \frac{\sqrt{2}\,(R+r_g)}{hsRr_g^2}, \text{ (approx.)} \tag{50}$$

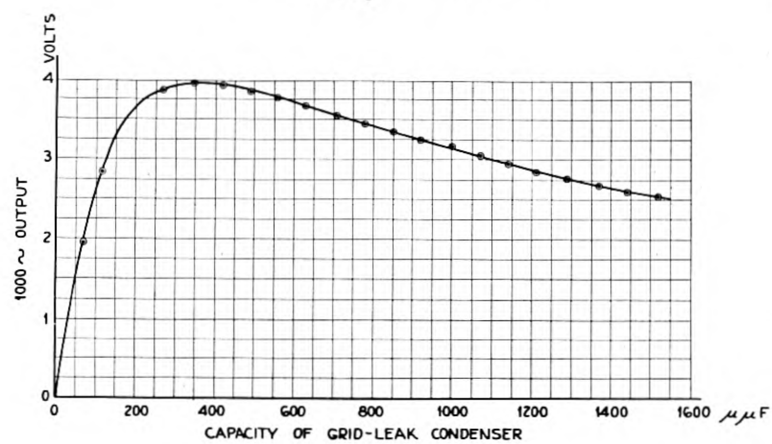


Fig. 6—Optimum size of grid-leak condenser

Experimental Conditions:

$h = 2\pi \times (30000 \pm 500)$

$s = 2\pi \times (1000)$

Grid-leak $= R = 10^6$ ohms

$r_g = 10^5$ ohms

Calculation Conditions:

$z_g \gg r_g$

$$q_h = \frac{i}{jhC}$$

$$q_s = \frac{\frac{R}{jsC}}{R + \frac{i}{jsC}}$$

Then the optimum size of the grid-leak condenser, C, is:

$$C^2_{opt} = \frac{\sqrt{2}\,(R+r_g)}{hsR\,r_g^2}$$

or;

$$C_{opt} = 361\mu\mu \text{ farad}$$

Fig. 6 illustrates the agreement between this relation and an actual circuit where the above conditions were closely approximated.

Plate Curvature Detection

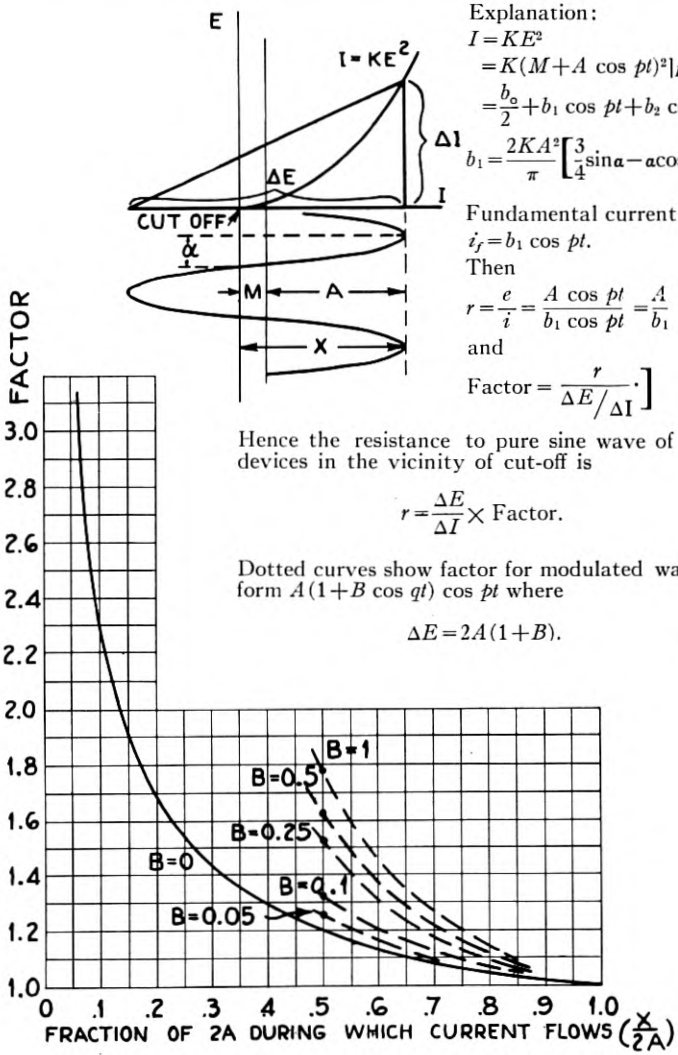
In discussing this phase of the problem we refer to equation (35). In addition to the remarks made in connection with that equation it is necessary only to add a few words on the evaluation of r_p and r_p'. In general, these quantities are susceptible to the same method of treatment that was suggested in dealing with r_g and r_g'. Two fundamental circuits for plate curvature detectors are in use. In the first the plate battery is placed in series with the load impedance. In the case when the load impedance contains appreciable resistance the normal or effective value of E_p must be obtained in the manner described for finding E_g. In the second circuit the plate battery potential is introduced through a low resistance, high impedance, choke, and the normal value of E_p is then equal to E_b. Especially in dealing with resistance coupled units these points should be borne in mind.

Amplification

Equation (33) gives the general amplification relation. The remarks made under the heading of the "Grid-Leak Detector" concerning the evaluation of the z's and q's are applicable here, as in all other vacuum tube relations. The special points to be brought out are the methods of applying the equations to so-called improper amplifiers of Class III. In this type of amplifier the grid swings negative further than the plate current cut-off point each cycle. Experience has shown that even in this event, to find the tube resistances, the approximation of using the secant line joining two points on the characteristic corresponding to the extreme values of the input voltage, is often justifiable. If greater accuracy is desired, the corrections given by the curve, Fig. 7, should be applied. These corrections are based on the assumption of a sine wave input and a characteristic that follows the square law, and to that extent are themselves in error. For modulated waves the dotted curves give values found by interpolation between the two points shown.

Modulation

The detection equations apply equally well to modulation effects. The only case in which a question may arise is that in which one of the input frequencies is introduced into the plate circuit of the tube



Explanation:

$$I = KE^2$$
$$= K(M + A \cos pt)^2]_{E+}$$
$$= \frac{b_0}{2} + b_1 \cos pt + b_2 \cos 2\,pt + \ldots$$
$$b_1 = \frac{2KA^2}{\pi}\left[\frac{3}{4}\sin\alpha - \alpha\cos\alpha + \frac{1}{12}\sin 3\alpha\right]$$

Fundamental current is:

$i_f = b_1 \cos pt.$

Then

$$r = \frac{e}{i} = \frac{A \cos pt}{b_1 \cos pt} = \frac{A}{b_1}$$

and

$$\text{Factor} = \frac{r}{\Delta E / \Delta I}.\Big]$$

Hence the resistance to pure sine wave of square-law devices in the vicinity of cut-off is

$$r = \frac{\Delta E}{\Delta I} \times \text{Factor.}$$

Dotted curves show factor for modulated waves of the form $A(1+B \cos qt) \cos pt$ where

$$\Delta E = 2A(1+B).$$

Fig. 7—Correction factor for resistance of non-linear device

while the other is introduced into the grid circuit. To analyze this condition for the general case, (see Fig. 8) let lower case e's refer to the driving voltage impressed directly on the grid. Let the E's refer to the driving voltage in series with an impedance in the plate circuit. We then have the series

$$i_p = a_1(E+e) + a_2(E+e)^2 + \ldots$$

which, in accordance with the complex quantity notation may be written

$$i_p = a_{1h}E + a_{1k}e + a_{2(2h)}E^2 + a_{2(2k)}e^2 + 2a_{2(OE)}E\bar{E} + 2a_{2(h+k)}Ee + 2a_{2(h-k)}E\bar{e} + 2a_{2(Oe)}e\bar{e} + \ldots$$

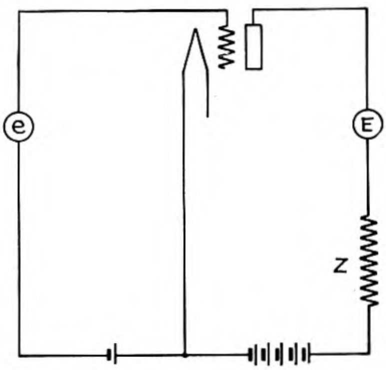


Fig. 8—Plate circuit modulation

Then, with the aid of (4), upon equating coefficients of like powers of e, E, and Ee, we get

$$a_{1h} = \frac{1}{r_p + z_h} \qquad a_{1k} = \frac{\mu}{r_p + z_k}$$

$$a_{2(2h)} = \frac{-\frac{1}{2} r_p r_p'}{(r_p + z_h)^2 (r_p + z_{2h})}$$

$$a_{2(2k)} = \frac{\frac{1}{2}\left[\frac{\partial \mu}{\partial E_g}(r_p + z_k)^2 + \mu \frac{\partial \mu}{\partial E_p}(r_p^2 - z_k^2) - \mu^2 r_p r_p'\right]}{(r_p + z_k)^2 (r_p + z_{2k})}$$

$$a_{2(OE)} = \frac{-\frac{1}{2} r_p r_p'}{(r_p + z_h)^2 (r_p + R)} \qquad a_{2(h+k)} = \frac{\frac{1}{2}\left[\frac{\partial \mu}{\partial E_p}\frac{r_p}{2}(2r_p + z_h + z_k) - \mu r_p r_p'\right]}{(r_p + z_h)(r_p + z_k)(r_p + z_{h+k})}$$

$$a_{2(h-k)} = \frac{\frac{1}{2}\left[\frac{\partial \mu}{\partial E_p}\frac{r_p}{2}(2r_p + z_h + \bar{z}_k) - \mu r_p r_p'\right]}{(r_p + z_h)(r_p + z_k)(r_p + z_{h-k})} \tag{51}$$

$$a_{2(Oe)} = \frac{\frac{1}{2}\left[\frac{\partial \mu}{\partial E_g}(r_p + z_k)^2 + \mu \frac{\partial \mu}{\partial E_p}(r_p^2 - z_k)^2 - \mu^2 r_p r_p'\right]}{(r_p + z_k)^2 (r_p + R)}$$

When z is a resistance, R, the expression for i_p reduces to

$$i_p=\frac{(\mu e+E)}{r_p+R}-\frac{\frac{1}{2}r_p r_p'}{(r_p+R)^3}E^2$$
$$+\frac{\frac{1}{2}\left[\frac{\partial\mu}{\partial E_g}(r_p+R)^2+\mu\frac{\partial\mu}{\partial E_p}(r_p^2-R^2)-\mu^2 r_p r_p'\right]}{(r_p+R)^3}e^2 \tag{52}$$
$$+\frac{\frac{\partial\mu}{\partial E_p}r_p(r_p+R)-\mu r_p r_p'}{(r_p+R)^3}Ee+ \text{ - - - .}$$

If μ is constant, this becomes

$$i_p=\frac{\mu e+E}{r_p+R}-\frac{\frac{1}{2}r_p r_p'}{(r_p+R)^3}(\mu e+E)^2+ \text{ - - -} \tag{53}$$

which shows that the circuit then acts as though a voltage, $(\mu e+E)$ had been impressed in series with the plate circuit.

Oscillation

The subject of vacuum tube oscillators has been so extensively treated elsewhere that but little new material has thus far been obtained from the general equations now offered. The method of handling the problem is, however, illuminating as it gives an example of what is meant by the statement that no sharply drawn line should be placed between oscillation, detection, amplification, or other uses of the thermionic vacuum tube.

In treating the oscillator problem we consider the amplification term of the general equations; namely

$$i_p=\frac{\mu e}{(r_p+z_n)}\frac{r_g}{\left[r_g+q_n(1-\frac{\mu}{\nu}\frac{z_n}{r_p+z_n}\right]}.$$

The oscillating conditions require that current shall flow without a driving voltage. Hence, as e is zero, i_p can be finite only if one of the factors in the denominator is zero. Thus either

$$r_p+z_n=0 \tag{54}$$

or

$$r_g+q_n(1-\frac{\mu}{\nu}\frac{z_n}{r_p+z_n}=0 \tag{55}$$

gives the conditions for oscillation. Fig. 5 and the relations of (45), (46), (47), (48) and (49) are applicable here. The condition of (54) requires a negative value of r_p, and hence is not the usual oscillation condition. The condition of (55) therefore gives the criterion for the oscillation condition. As before, neglecting quantities in $\frac{1}{\nu}$ we may write (55) in the following form

$$r_g + q_n = 0$$

or

$$r_g + \frac{z_1(z_g + r_g)}{z_g} = 0. \tag{56}$$

When applied to a hypothetical Hartley oscillator, Fig. 9, with the circuit constants

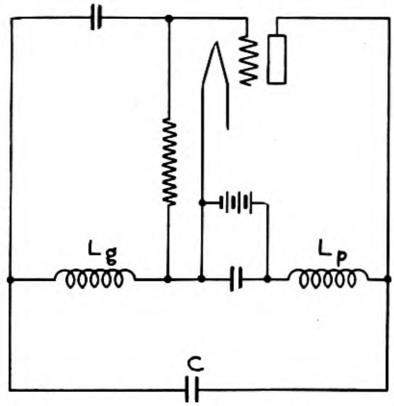


Fig. 9—Hartley oscillator

$$z_2 = j\omega L_p \qquad z_3 = \frac{1}{j\omega C} \qquad z_1 = j\omega L_g,$$

equation (56) gives as the conditions for oscillation

$$\omega^2 = \frac{1}{\left[L_p + L_g + \frac{L_p L_g}{C r_p r_g}\right]} = \frac{1}{(L_p + L_g)\,C} \text{ (nearly)}, \tag{57}$$

$$L_p = \left[\mu - \frac{r_p}{r_g L_p}\right] L_g = \mu L_g \text{ (nearly)}. \tag{58}$$

The relations of (57) and (58) have been given many times, and are included here only in order to illustrate the ease with which simple problems may be solved from fundamental relations.

Application of the Theory

The illustrations will serve to give a sufficiently comprehensive view of the methods of applying the general equations to special cases.

Inasmuch as the derivation of the equations requires no assumptions other than that the static curves of grid current-grid potential, and plate current-plate potential of the tube remain constant, the accuracy with which a given problem may be calculated depends only upon the ability to determine the effective differential coefficients required by the Taylor's series expansions, and the number of terms of the series included. Practically, the component of current of a given frequency resulting from any higher order term is entirely negligible with respect to the component of the same frequency resulting from lower order terms. For precise results in a general case the calculations are necessarily tedious, since the physical processes are quite complex. However, in any given special case one of the respective approximations indicated is usually allowable, which greatly simplifies matters. In the event that any question arises concerning the proper phase angles for the complex impedances, the correct result may always be arrived at by writing the voltages in full complex form, as illustrated in the mathematical digression. The impedances will then take care of themselves.

While it is difficult to show mathematically the convergence of the series of (31), experience has shown that the convergence is so rapid that higher order terms may be neglected, unless new frequencies developed by them are under investigation. In these cases, the conditions of the problem are often such that simplifying assumptions may be made at the outset. If familiarity with the complex impedances has been attained, it will, in many cases, be sufficient to derive all equations on the basis of resistance only, and then introduce the complex impedances in the manner indicated by the analogy between these and the general equations.

The higher order coefficients are given below for the special case where resistances, only, are considered, and where the voltage, e_g, is known. It is found more convenient to use the P's, equation (4), in their derivative form than to attempt to express them in terms of μ and r_p, so referring to the expansion

$$i_p = a_1 e_g + a_2 e_g^2 + a_3 e_g^3 + a_4 e_g^4 + a_5 e_g^5 + \cdots,$$

we have

$$a_1 = \frac{P_1}{1+P_2Z}$$

$$a_2 = \frac{\frac{1}{\underline{|2}}}{1+P_2Z}\left[P_3 - 2P_4\frac{P_1Z}{1+P_2Z} + P_5\frac{Z^2P_1^2}{(1+P_2Z)^2}\right],$$

$$a_3 = \frac{1}{1+P_2Z}\left(\frac{1}{\underline{|2}}\left[-2P_4a_2Z + 2P_5a_1a_2Z_1Z_2\right]\right.$$

$$\left. + \frac{1}{\underline{|3}}\left[P_6 - 3P_7a_1Z + 3P_8a_1^2Z^2 - P_9a_1^3Z^3\right]\right),$$

$$a_4 = \frac{1}{1+P_2Z}\left(\frac{1}{\underline{|2}}\left[-2P_4a_3Z + P_5(a_2^2Z^2 + 2a_1a_3Z_1Z_3\right]\right.$$

$$+ \frac{1}{\underline{|3}}\left[-3P_7a_2Z + 6P_8a_1a_2Z^2 - 3P_9a_1^2a_2Z^3\right]$$

$$\left. + \frac{1}{\underline{|4}}\left[P_{10} - 4P_{11}a_1Z + 6P_{12}a_1^2Z^2 - 4P_{13}a_1^3Z^3 + P_{14}a_1^4Z^4\right]\right)$$

$$a_5 = \frac{1}{1+P_2Z}\left(\frac{1}{\underline{|2}}\left[-2P_4a_4Z + 2P_5(a_1a_4Z^2 + a_2a_3Z^2)\right]\right.$$

$$+ \frac{1}{\underline{|3}}\left[-3P_7a_3Z + 3P_8(a_2^2Z^2 + 2a_1a_3Z^2)\right.$$

$$\left. - P_9(a_1Z\overline{a_2^2Z^2 + 2a_2a_3Z^2} + 2a_1a_2^2Z^3 + a_1^2a_3Z^3)\right]$$

$$+ \frac{1}{\underline{|4}}\left[-4P_{11}a_2Z + 12P_{12}a_1a_2Z^2 - 12P_{13}a_1^2a_2Z^3 + 4P_{14}a_1^3a_2Z^4\right]$$

$$+ \frac{1}{\underline{|5}}\left[P_{15} - 5P_{16}a_1Z + 10P_{17}a_1^2Z^2 - 10\,P_{18}a_1^3Z^3\right.$$

$$\left.\left. + 5P_{19}a_1^4Z^4 - P_{20}a_1^5Z^5\right)\right],$$

APPENDIX I

To Show that with Negative Grid Potentials the Relation:

$$\mu \frac{\partial \mu}{\partial E_p} = \frac{\partial \mu}{\partial E_g}$$

Holds With Fair Precision

We have the fundamental expression:

$$I_p = I_p(E_g, E_g) \tag{1}$$

Suppose that E_g and E_p are allowed to vary under the restriction that I_p is maintained constant. Then:

$$d\ I_p = 0 \tag{2}$$

Hence:

$$\frac{dI_p}{dE_g} = 0 = \frac{\partial I_p}{\partial E_g} + \frac{\partial I_p}{\partial E_p} \frac{dE_p}{dE_g} \tag{3}$$

Whence:

$$\left. \frac{dE_p}{dE_g} \right]_{I_p} = -\mu \tag{4}$$

Also:

$$d^2 I_p = 0 \tag{5}$$

Hence:

$$\frac{d^2 I_p}{dE_g^2} = 0 = \frac{\partial^2 I_p}{\partial E_g^2} + 2 \frac{\partial^2 I_p}{\partial E_g \partial E_p} \frac{dE_p}{dE_g} + \frac{\partial^2 I_p}{\partial E_g^2} \left(\frac{dE_p}{dE_g} \right)^2 + \frac{\partial I_p}{\partial E_p} \frac{d^2 E_p}{dE_g^2} \tag{6}$$

Then with the aid of (4), above, and (6) in the body of the paper, we get

$$\frac{\partial \mu}{\partial E_g} - \mu \frac{\partial \mu}{\partial E_p} + \frac{d^2 E_p}{dE_g^2} = 0 \tag{7}$$

Equation (7) shows that:

$$\frac{\partial \mu}{\partial E_g} = \mu \frac{\partial \mu}{\partial E_p} \tag{8}$$

provided that:

$$\frac{d^2 E_p}{dE_g^2} = 0 \tag{9}$$

when I_p is constant.

Experimental curves showing the relation between E_p and E_g required to maintain I_p constant are straight lines, to a very close approximation, in the region where the grid potential is negative with respect to the filament as shown in Fig. 10. Hence, in this region (9) is satisfied for all practical purposes, and, therefore, the proof of (8) follows directly.

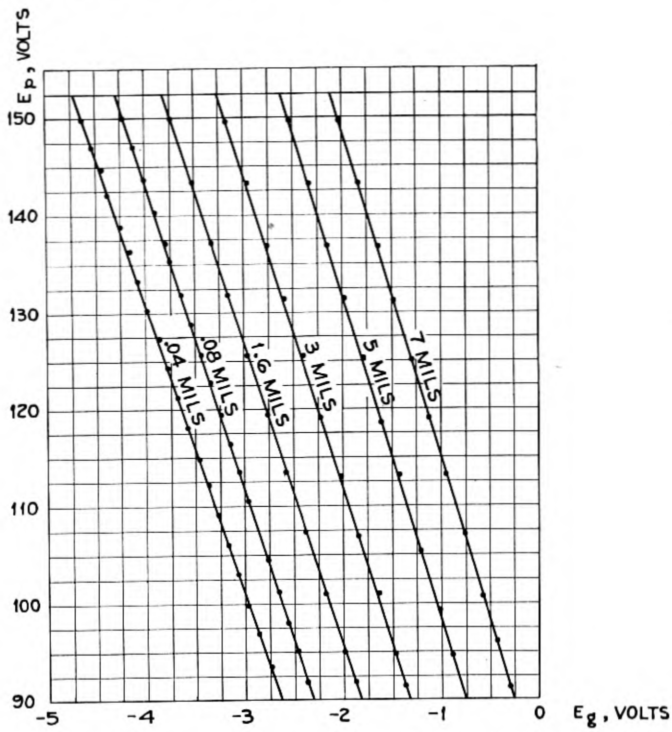


Fig. 10—Relation between E and E for constant plate current

Contemporary Advances in Physics—XI
Ionization

By KARL K. DARROW

IONIZATION, in its most general sense, signifies a segregation of positive from negative charge within the volume of a substance which as a whole is (or initially was) electrically neutral. In practice a gas (for instance) is said to be *ionized* if charges of either sign can be extracted from it. Charged particles of both signs, electrons and ions, can be drawn out from a gas in which an electrical discharge is being maintained; in such a condition, therefore, a gas is ionized. Millikan's droplets, floating around in a gas which had recently been irradiated, absorbed charges of either sign out of the gas, which therefore was ionized by the radiation and remained ionized for some time afterward. A negatively-charged electrode immersed in a carefully-screened gas receives very little charge from it; this condition continues if the gas is bombarded with electrons having less than a certain speed; let the speed of the bombarding electrons be increased past this limit, and the electrode begins to receive positive charge—the gas is ionized by the electrons. Dilute electrolytic solutions are evidently in a continual and spontaneous state of ionization.

Observations on positive ions issuing from ionized gases have been interpreted as meaning that all such ions are atoms or molecules bearing charges of which the magnitude is e, or $2e$, or some other small-integer multiple of e; in other words, as meaning that positive ions are atoms or molecules from which one or more electrons have been detached. Generalizing from these to all cases, it is believed that the first stage of ionization, in monatomic gases at least, is the detachment of electrons from atoms. Whether the separated electrons remain free, or attach themselves to other atoms, or become the gathering-agents of clusters of atoms, is an interesting but at present subsidiary question. Ionization in monatomic gases begins by the detachment of electrons from atoms; and the word "ionization" in fact is frequently used to mean this process alone. In diatomic and compounded gases, the nature of the ions observed permits either of two suppositions; the initial process of ionization may be the detachment of electrons from molecules, or the splitting of molecules into fragments each consisting of one or more atoms, some of these fragments having an excess and the others a compensating deficit of electrons. Special experiments must be performed to decide between these suppositions.

While self-sustaining discharges in gases may produce a vast variety of identifiable ions, they are not suitable for revealing the process of producing these ions. By bombarding a gas with electrons of known speed, ions may be produced under very simple and intelligible conditions. It is then found that in order to detach an electron from an atom of a monatomic gas, a definite amount of energy, the *ionizing-energy* of the gas, must be transferred to the atom. The ionizing-energy is not unique; for most kinds of atoms there are several distinct quantities answering to the same definition. Nevertheless there is one particular and outstanding value which is particularly known as *the* ionizing-energy or *ionizing-potential.* It varies periodically from element to element along the Periodic Table, and is therefore ascribed to an outer electron of the atom; indeed it may be described as the *extraction-energy for the outermost or loosest electron.*

Of the other values of ionizing-energy for a given atom, some are lower than the principal ionizing-potential. These, however, are attributed to atoms in abnormal states. The others are greater than the principal ionizing-potential; some of them are very much greater and increase steadily from one element to the next along the Periodic Table, and are therefore ascribed to deeper-lying electrons and may be described as *extraction-energies for inner electrons.*

The spontaneous ionization of radioactive substances is an entirely irregular function of atomic number and is attributed, for this and other reasons, to events occurring in the nuclei.

At this point it is necessary to define some units. In most determinations of ionizing-energies, a stream of electrons originally moving with speeds thought negligibly small is accelerated by a potential-rise and then projected into the gas under examination. Their kinetic energies in ergs are thus given in terms of the *voltage* V of the potential-rise by the equation

$$\text{Kinetic Energy} = eV/300 = 1.591 \cdot 10^{-12}\ V. \qquad (1)$$

It is customary to measure the kinetic energy of an electron by the voltage-rise which gave it, or could have given it, that energy; which is tantamount to employing a unit of energy equal to $1.591.10^{-12}$ erg. This unit may be called the *equivalent volt.*

$$\textit{One equivalent volt} = 1.591 \cdot 10^{-12}\ \textit{erg} \qquad (2)$$

The name, it must be admitted, is neither short nor elegant; at all events it is preferable to the slovenly usage of speaking of an electron as having so many "volts of energy" (!) or a "speed of so many volts" (!!) On the other hand, it seems quite unobjectionable to speak of an electron having a kinetic energy of one equivalent volt as a "one-volt electron."

The ionizing-energy of an atom is usually given in equivalent volts, whence the name *ionizing-potential.*

Occasionally one meets with a value stated for an ionizing-potential in terms of a unit known as the *wave-number* ("equivalent wave-number" would be better) which amounts to $1.968.10^{-16}$ erg.

Ionization-Potentials

The ionizing-potential of a monatomic gas is usually measured by projecting electrons with controllable kinetic energy K into the gas, and determining the value of K at which current begins to flow into an electrode inserted into the gas and maintained at such a potential that positive ions, but no electrons, can reach it.

This method requires more elaborate apparatus than the outline suggests. The experimenter must guard against an effect which was not suspected by those who first worked with the method. Electrons having kinetic energy less the ionizing-energy of the gas may cause the atoms which they strike to emit radiation. Some of this radiation falls upon the electrode arranged to collect positive ions, and expels electrons from it. The field around the collecting-electrode, being such as to draw positive ions toward it, drives these electrons away; and so there is a continuous current of negative charge out of the electrode into the gas, which is quite indistinguishable from a current of positive charge out of the gas into the electrode. Thus the value of K at which positive charge first seems to flow into the electrode from the gas is the "critical" electron-energy (as the phrase is) not for producing ionization but for producing radiation. The earliest determinations of what were thought to be ionizing-potentials were vitiated by this effect.

To avoid or recognize the influence of radiation several schemes have been devised.[1] For example, if two collecting-electrodes are used in alternation, one having a large area and the other being small, much more radiation will fall upon the larger one, and there will be a correspondingly great difference between the currents of negative charge out of the two; but if ions are being formed in the gas, the difference between the numbers of these which find their way to the large and to the small electrode will be much less pronounced. A slender collecting-electrode may record only a very small current due to radiation, but a

[1] For a detailed account of the methods developed up to 1924, consult K. T. Compton and F. L. Mohler: "Critical potentials" (*Bull. Nat. Res. Council,* No. 48).

very large one whenever ionization commences. This scheme has been adopted by K. T. Compton.

Another, and the most common, device for distinguishing ionization from radiation consists in surrounding the collector with a sheath of metal gauze, maintained at a potential slightly (say 3 volts) more negative than the electrode which it screens. Positive ions pass through its meshes to the collector, somewhat slowed down but not driven back. Radiation also passes through the meshes to the collector, but the electrons which it drives out are turned back by the adverse field and reenter the metal whence they came, so that the net result is the same as though they had never come out. Ionization thus produces a current of positive charge into the collector, and radiation none; or radiation may even produce a current of negative charge into the collector, thus accentuating the contrast, for electrons which are ejected from the metal gauze are drawn to the screened electrode. This is the scheme devised by F. S. Goucher.

Another, and possibly the best, method for measuring ionizing-potentials is quite insensitive to radiation. A hot filament is immersed in the gas, which may be supposed to be surrounded by connected metal walls so that its boundaries are all at the same potential. If the efflux of electrons from the filament is so plentiful that it is limited by space-charge,[2] and this condition persists as the potential-difference between walls and filament is raised to the value just sufficing to give to the electrons energy enough to ionize the gas, then at the moment of incipient ionization the space-charge limitation is partially or totally cancelled, and the current increases sharply. This is I. Langmuir's method. It is better to keep the potential-difference between the walls and the filament small and constant, and admit into the gas electrons with controllable energy from another source; when the energy of these auxiliary electrons is raised to attain the ionizing-potential of the gas, the current from the filament suddenly increases. This is the method of G. Hertz[3] and K. H. Kingdon.[4]

Most of the accurate measurements of ionizing-potentials have been made with a collecting-electrode sheathed by a gauze, according to the precept of Goucher. The apparatus is a complicated affair, for the parts already mentioned are by no means all that are required; in some cases the whole interior of the tube appears to be webbed with gauzes. A hot filament (occasionally an illuminated metal plate) is provided as source for electrons, and its potential—or the potential of

[2] See the sixth article of this series (December, 1924).

[3] *ZS. f. Phys.* 18, pp. 307–316 (1923).

[4] *Phys. Rev.* (2) 21, pp. 404–418 (1923).

its negative end—is taken as the zero from which the other potentials are measured. In the sketch (Fig. 1) this is marked F. Close to the source there is a gauze (G_1) maintained at the controllable potential V and thus providing the potential-rise by which the electrons are accelerated. It is clearly desirable that the electrons should move at their

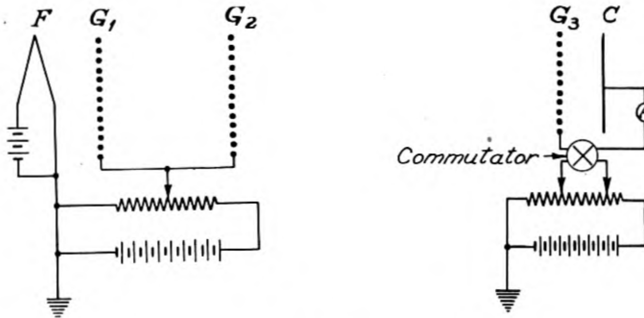


Fig. 1

known maximum speed over as long a path as possible in the gas; consequently a second gauze (G_2) is set up beyond G_1, and maintained at nearly the potential V so that there is a nearly equipotential region between them. (Generally the potential of G_2 is raised a fraction of a volt above V so that there may be a slight impulsion of the ions formed between G_1 and G_2 toward the collector.) Beyond G_2 are the collector C and its protecting gauze G_3, maintained at potentials lower than the filament so that no electrons may reach them. The current of which the sign indicates whether it is due to radiation or ionization, as was explained above, flows through the galvanometer at A.

With such an apparatus as this it seems to be easy enough to measure ionizing-potentials correctly within one or two volts. As soon as greater accuracy is sought after, the real troubles begin. The electrons do not all leave the source with negligible speed; their speeds are distributed over a finite range. The potential to which they climb in passing through the meshes of a gauze is not quite equal to the potential of the gauze-wires themselves. The potential-differences between the different electrodes are not accurately given by voltmeters, for there are contact-potential-differences superposed upon the values indicated. The filament is not an equipotential surface if it is heated by a current, although this difficulty can be overcome if the experimenter thinks it worth the trouble. Electric charges marooned upon the walls of the tube, electrons ejected by radiation from the gauze G_3 and accelerated backwards to G_2 with a final speed higher than the electrons from F

ever attain, are capable of causing false conclusions. The third significant figure in the value of an ionizing-potential is many times harder to attain than the first two; and it is not surprising that many experimenters have chosen to mix some standard gas such as helium into the gases with which they experimented, and to determine the difference between the ionizing-potentials of the standard gas and the other gases, rather than any of them absolutely.

Before bringing out the numerical values of ionizing-potentials, I must allude to the fact that the quantity measured in these experiments is the kinetic energy possessed by the electrons when they are just able to ionize the atoms, which might not be the same thing as the energy actually transferred to the atoms. A particle of mass m moving with speed u has not only kinetic energy $K = \frac{1}{2}mu^2$ but also momentum mu. If it impinges against a previously-stationary particle of mass M, and momentum is conserved in the impact, then the particles must be in motion after the impact, and some of the initial kinetic energy of the striking particle must be saved, so to speak, to provide for this motion. What is left over is available for ionization or other purposes. Without involving ourselves in the general case, we may note that the most favourable conceivable case for having a large proportion of energy left over, when the striking particle is less massive than the struck one, is that in which the more massive particle has all the momentum after the impact. Suppose therefore that after the impact the striking electron and the liberated electron are both stationary, and the ion of mass M is moving with speed V. Conservation of momentum is expressed by writing:

$$mu = MV. \tag{3}$$

The energy T available for ionization or other purposes is given by:

$$K = \tfrac{1}{2}mu^2 = \tfrac{1}{2}MV^2 + T. \tag{4}$$

so that

$$T = K(1 - m/M). \tag{5}$$

Since the masses of atoms range from 1845 to nearly half a million times the mass of an electron, an electron might spend over 999 promille of its energy in ionizing an atom; and therefore there is no essential impossibility in supposing that the energy possessed by an electron just able to ionize is actually equal, within the uncertainty of measurement, to the ionizing-energy of the atom. This supposition is confirmed by the agreements between observed ionizing-potentials and the theoretical values deduced from spectra by using Bohr's method of interpretation.

If the gas or the electron-stream is extremely dense, positive ions appear when the energy of the bombarding electrons is lower than the ionizing-energy as determined by experiments with more rarefied gas or a scantier stream of electrons. Various reasons are assigned for this in various cases; one fundamental reason is, that an atom struck by an electron having less than the ionizing-energy may be put into abnormal states of some duration, in which it can be ionized by receiving a smaller amount of energy than would ionize it in its normal state.

In Fig. 2 the measured values of ionizing-potential are plotted.[5]

There is a way of expressing these and other yet-to-be-presented facts about ionizing-energies, which at this point will probably seem

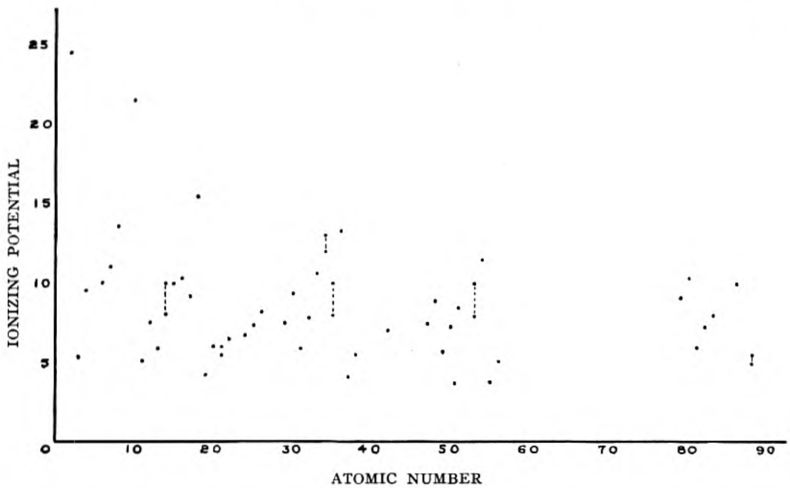


Fig. 2

unnatural but later will be highly convenient. Suppose that by transfer of the ionizing-energy V_0 to an atom it is converted into a system composed of an ion bearing charge $+ e$ and a free electron. This system has potential energy V_0 relatively to the normal state of the atom. The detached electron may wander off and the ion eventually unite itself with another electron. It is convenient, therefore (whether or not it is strictly legitimate) to think of this potential energy V_0 as being associated with the ion alone; and to say that the atom possesses, in addition to its normal state, one or more *states of ionization* or *states of the ionized atom*, each of them characterized by a certain value of po-

[5] I am deeply indebted to Professor F. A. Saunders, who has kept a current catalogue of published values of ionizing-potentials, for enabling me to copy his tabulations.

tential energy. The ionizing-potential of the atom is then, by definition, equal to the potential energy of that state of ionization which differs least in energy from the normal state.

Another way of describing the ionizing-potential is to say that it is the *energy required to detach the loosest electron from the atom.* This involves a picture of an atom as a system of separately-identifiable electrons, "bound" with various degrees of looseness or tightness. Such a picture is so nearly indispensable, that there need be little hesitation about introducing it here. Frequently the term *valence-electron* is used instead of "loosest electron"; there is little in its favour beyond the general inability of physicists to think of a better one.[6]

Detachment of the Loosest Electron by Other Agencies than Electron-Impacts

Other agencies than the blows of electrons are capable of detaching the loosest electron from an atom; but it is very much more difficult to obtain simple and intelligible information about their immediate effects than about those of electron-impacts.

The study of *ionization by radiation* involves a host of new problems. Theoretically the conditions seem simple enough. Radiation of any frequency ν behaves in some respects as though it consisted of streams of particles each having energy $h\nu$ and momentum $h\nu/c$. Since it behaves in this manner in so far as absorption in gases and ejection of electrons from solids are concerned, we should expect it to do likewise in effecting ionization of atoms. If so, radiation should ionize atoms if and only if its frequency ν equals or exceeds a critical or threshold value ν_0, expressed in terms of the ionizing-potentials V_0 of the atoms (measured in equivalent volts) by

$$h\nu_o = eV_o/300. \tag{6}$$

Projecting light from a spectrum upon a gas, and passing steadily from low to high values of ν, we should expect ionization to commence abruptly at ν_0.

Experimentally, the task of testing this inference has baffled everyone, at least until very recently. In the first place, the values of threshold-frequency ν_0 for various atoms correspond to values of threshold-

[6] The terms "optical electrons" and "series electrons" are sometimes seen; they are derived from theoretical pictures which are in danger of mutation (some people now ascribe most series-spectra to displacements of electrons in groups). The German term "Leuchtelektron" probably sounds better in German than its equivalent "shining electron" would sound in English. It may be remembered that difficulty in choosing a good name for a concept sometimes signifies that the concept is essentially vague and not rooted in Nature.

length λ_0 lying between 504A (helium) and 3184A (caesium); and this is the most troublesome region of the spectrum to deal with, partly because light of wavelengths lying within it is tremendously absorbed by nearly all solids and even gases, and partly because good sources for such light are difficult or impossible to procure. Even in the comparatively accessible zone between 2000A and 3500A it is customary to use the light of the mercury arc, which provides a few widely-spaced bright spectrum-lines; as though in determining ionizing-potentials by electron-impacts one had to use electrons of certain distinct and widely-spaced energy-values, and could not refine the measurements by adjusting the accelerating voltage to intermediate values *ad libitum*. In measuring ionizing-potentials by electron-impacts there is a secondary difficulty due to radiation from struck atoms falling upon the collector; here the difficulty becomes a primary one, since the primary radiation itself is competent to produce this effect. The effect is most vicious with alkali-metal vapours, as they deposit themselves over all the solid surfaces of the apparatus in films excessively liable to pour out electrons when stimulated by light or warmth; yet these are the only elements for which λ_0 lies above 2500A.

Several experimenters have minimized the undesired effects of the radiation by projecting a narrow beam of light across a jet of alkali-metal vapor boiling up out of a narrow channel in the main tube. The beam struck nothing except the jet and beyond it a "trap" in which presumably it was totally absorbed and no part was scattered. The jet passed onward, near to an electrode negatively charged to receive positive ions. With potassium vapors, for which λ_0 should be 2856A, R. C. Williamson found ionization commencing somewhere between 3100A and 2800A; H. Samuel thought that it commences between 2804A and 2893A; E. Lawrence concluded that it begins at 2610A.[7] P. D. Foote and F. L. Mohler[8] detected the positive ions by their effect in annulling the space-charge limitations upon the current from a hot filament, after the fashion of the last-mentioned method of determining ionization-potentials. Their result was somewhat unexpected; they found ionization in caesium vapor at wavelengths even greater than the threshold-wavelength. This is attributed to the same cause as brings about a lowering of the apparent ionizing-potential when dense

[7] *Phys. Rev.* (2) 27, pp. 37–51 (1926); 26, pp. 197–207 (1925).

[8] E. O. Lawrence, *Phil. Mag.* 50, pp. 345–359 (1925); R. C. Williamson, *Phys. Rev.* 21, pp. 107 (1923); H. Samuel, *ZS. f. Phys.* 29, pp. 209–213 (1924); and prior literature cited in the first two. In all of the cited experiments the vapor had freshly issued from condensed potassium, and may have contained a large proportion of molecular aggregates, to which Lawrence attributes the difference between his observed threshold-wavelength and the calculated ν_0, Cf. also G. F. Rouse and G. W. Giddings, *Proc. Nat. Acad. Sci.* 11, pp. 514–177 (1925).

streams of bombarding electrons are used: that is to say, it occurs because light of less than the threshold frequency puts some of the atoms into abnormal states, in which less energy is required to ionize them than in the normal state.

The study of *ionization by positive ions* is also very troublesome. This is partly because there are no such convenient sources for controllable positive ions as there are for electrons. The ions emerging from hot filaments are generally not all of one kind. If ions of a particular sort, hydrogen ions for instance (these would give the most valuable information of any) are produced by bombarding the proper kind of gas by electrons having a suitable ionizing-energy, they cannot be used for ionizing except in the same tube and therefore upon the same gas; further, it is necessary to keep the bombarding electrons out of the region where the positive ions are meant to ionize, by an elaborate system of gauzes and opposing potentials. If the collecting electrode is maintained at a positive potential so as to receive electrons produced by the ionization, it receives also the electrons which are knocked out of the walls of the tube by positive ions which strike them. It is scarcely surprising, then, that the published data are scanty and not always concordant.[9]

The considerations about conservation of momentum during impacts, mentioned in dealing with ionization by electrons, show that we should hardly expect a positive ion to be able to ionize unless it has much more energy than must be transferred to the atom to detach the loosest electron from it; twice as much, if the ion is of the same mass as the atom.

Identification of Ions Produced by Electron-Impacts

The methods hitherto described for detecting the onset of ionization in a gas show when free positive charges appear in a gas, but give no further information about them. The methods employed by J. J. Thomson and F. W. Aston reveal the charge-to-mass ratios of ions occurring in a gas carrying a self-maintaining discharge, but give very little information about the precise conditions necessary to produce them. A combination of methods of these two kinds was first effected by H. D. Smyth.[10]

One of the tubes employed by Smyth is sketched in Fig. 3. Electrons from the filament F are accelerated through the potential-rise V_1 to the

[9] For work published up to 1922 see the review and bibliography by A. J. Saxton, *Phil. Mag.* 44, pp. 809–823 (1922). See also J. T. Tate, *Phys. Rev.* (2) 23, pp. 293–294 (1924).

[10] *Proc. Roy. Soc.* A102, pp. 283–293 (1922–23); A104, pp. 121–134 (1923); *Phys. Rev.* (2) 25, pp. 452–468 (1925) and references there given.

gauze E_1, and then turned back by an adverse potential-fall V_2 before they reach the partition E_2 pierced by the slit S_2. Positive ions produced by the electrons in the region between E_1 and E_2 are drawn

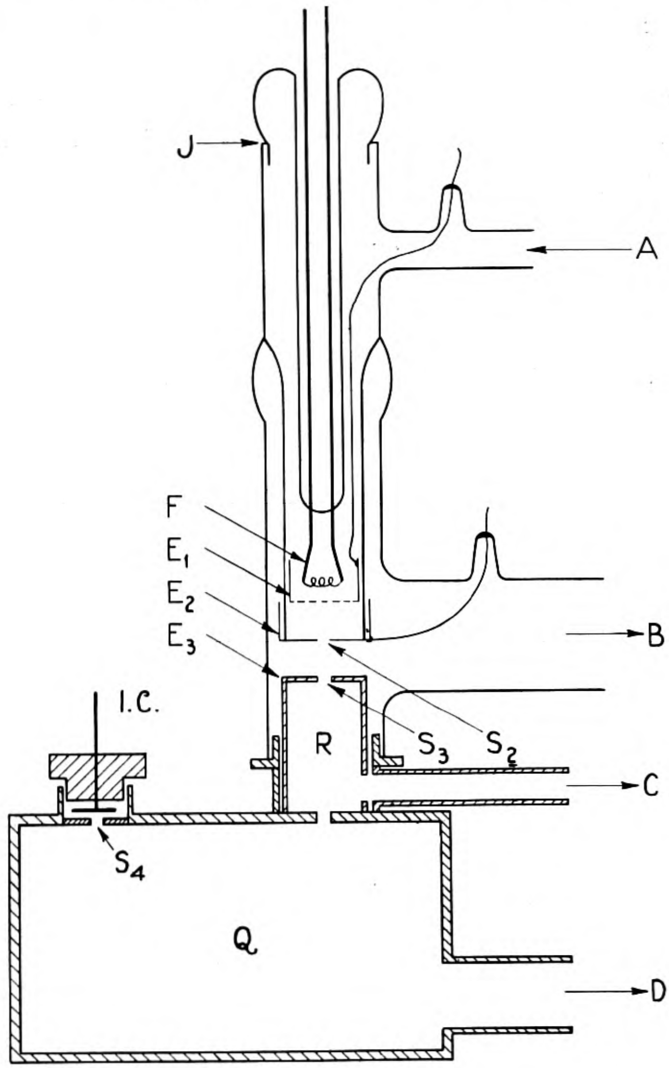


Fig. 3

toward E_2; some of them emerge through S_2, and encounter an additional potential-fall V_3 which draws them to the partition E_3. Those which pass through the slit S_3 are now ready, after passing through the field-free region R, to be swung around in semi-circular arcs by a magnetic field H applied normally to the plane of the paper over the region Q; thus they arrive at the ion-collector behind the slit S_4. The major experimental difficulty consists in maintaining simultaneously a gas-density between F and E_2 high enough to afford plenty of ions, and a gas-density in R and Q low enough so that the ion-stream is not dispersed. This is effected by feeding in the gas through A and applying powerful pumps to draw it out through B, C and D.

Varying H and plotting against it the current into the ion-collector, one obtains a curve with peaks, such as the one in Fig. 4. This is, how-

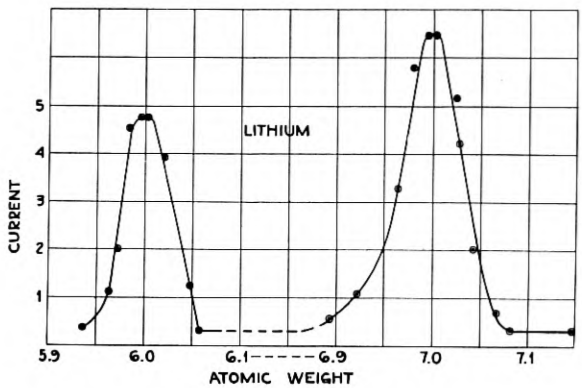


Fig. 4

ever, a curve obtained by Dempster with ions issuing from a hot filament. The charge-to-mass ratio for the kind of ion producing each peak is calculated from the accelerating-voltages, the deflecting field, and the diameter of the circular arc through which they swing.

The use of this method in determining ionizing-potentials may be illustrated from the work of H. A. Barton on argon.[11] Observing at values of V_1 superior to some 50 volts a two-peaked curve with the M/E values of the corresponding ions standing in the ratio 2:1; and observing at values of V_1 inferior to some 40 volts only one of these peaks, the one with the greater value of M/E; he inferred that this peak was due to A^+ ions and the other to A^{++} ions. Plotting the heights of these peaks or the areas under them as functions of V_1 he obtained curves such as those shown in Fig. 5. From many such curves as these he deduced that the energy of electrons just able to produce doubly-ionized argon atoms exceeds that of electrons just able to produce

[11] *Phys. Rev.* (2) 25, pp. 469–483 (1925).

singly-ionized argon atoms by 30 equivalent volts.[12] In the same manner, Smyth concluded that the energy of electrons just able to produce doubly-charged mercury ions exceeds by about 9 equivalent volts that of electrons just able to produce singly-charged mercury ions. The method, however, has been used chiefly for studying diatomic gases, and therefore will be mentioned in another section.

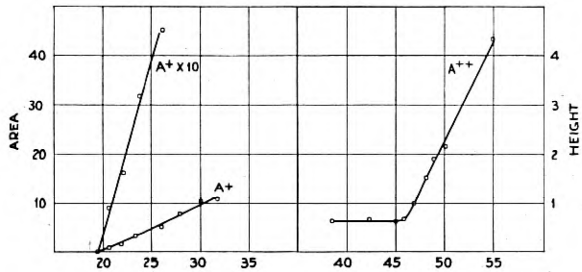


Fig. 5

Ionization of Molecular Gases [13]

The experiments of Thomson and Aston upon the ions proceeding from self-sustaining discharges in molecular gases show that these comprise individual atoms and also molecules of various sorts, each deprived of one or occasionally of more than one electron. Not all of these, however, are produced by the direct and simple agency of a single electron-impact against a normal molecule; some of them result from encounters of ions originally produced in the discharge with molecules which they meet in the gas, either in that region where the discharge is being maintained or in the channel through which they pass to reach the analyzing fields. This stands out very clearly in such experiments as one performed by A. J. Dempster, who projected 800-volt electrons into hydrogen gas and determined the relative abundance of the ions H^+, H_2^+ and H_3^+ arriving at his collecting-electrode after passing through a certain distance in the gas. At a gas-pressure amounting to .01 mm. Hg, the H_3^+ ion was the most plentiful of all and the other two not far behind; at .0017 mm. Hg both the H^+ and H_3^+ ions were definitely less abundant than H_2^+, and below .0005 mm. the H_2^+ ion

[12] Actually he obtained 17.3 volts for the one critical potential, 47.4 for the other, and assumed that the difference between 17.3 and the accepted value of 15.2 for the first ionizing-potential of argon is due to contact potentials and other influences affecting each of the observed critical potentials equally.

[13] For a general bibliography of this subject see T. R. Hogness & E. G. Lunn, *Phys. Rev.* (2) 26, pp. 44–55, 786–793 (1925); also V. Kondratjeff, *ZS. f. Phys.* 22, pp. 1–8 (1924) and 31, pp. 535–541 (1925).

was left almost alone upon the scene. These results signify that an 800-volt electron operates ionization in hydrogen by detaching an electron from a molecule; other kinds of ions appearing in the gas are due to subsequent adventures of these ions.

The method of H. D. Smyth is suitable for investigations into this question. In apparatus such as his, hydrogen bombarded by (say) 40-volt electrons is found to contain all three ions H^+, H_2^+ and H_3^+; but as the density of hydrogen is reduced, the first and the last of these ions become less abundant and finally insignificant by comparison with the ion H_2^+. As the bombarding-voltage is reduced towards the value (about 16) at which ionization commences, all three kinds of ions become less plentiful; but with high densities and sufficiently sensitive apparatus it is found that H_3^+ makes its appearance as early as H_2^+, and there is no reason not to suppose the same about H^+. In hydrogen, therefore, and also in nitrogen, it is agreed that an electron-impact against a molecule results, if in any sort of ionization at all, in the detachment of an electron from the molecule, not (for instance) in a dissociation into one ionized atom and another atom ionized or neutral. Dissociation and new sorts of association may result from the further adventures of this molecule-ion in the gas. In certain compound gases[14] of which the molecules consist of two or more atoms of different kinds, there is reason to expect the contrary: that is, that an electron-impact against a molecule would result directly in splitting it into a positively-charged atom (or group of atoms) and a negatively-charged atom (or group of atoms). Certain experiments indicate this: in $ZnCl_2$ vapor, for instance, Cl atoms bearing an extra electron and ZnCl molecules minus an electron are found as soon as ionization commences; but the question can hardly be deemed settled until comparative measurements are made at various gas-densities.

From these experiments it follows that a measurement of the energy just sufficient to produce ions in a molecular gas, while interesting in itself, can hardly be interpreted without additional data regarding the nature of the ions produced. There are other difficulties in determining ionizing-potentials in such gases; for instance the likelihood that the hot filament will itself dissociate the gas. The published determinations are frequently contradictory; the various published values for the ionizing-potentials of hydrogen, for instance, form one of the most discouraging sets of irreconcilable data to be found in physics.

According to thermochemical measurements the "heat of dissociation" of hydrogen, in other words the energy-difference between a system of two free H atoms and an H_2 molecule, amounts to 3.5 equivalent

[14] Those designated by chemists as heteropolar.

volts. One would expect to be able to dissociate hydrogen by bombarding the gas with 3.5 volt electrons; yet nothing of the sort happens. This is an instance of the frequently-occurring observation that a particle or a quantum may have abundant energy to produce a particular effect and yet be quite unable to produce it. One would expect also that the minimum energy required to convert an H_2 molecule into an H^+ ion and an H atom and a free electron would exceed by 3.5 equivalent volts the ionizing-energy of an H atom, yet the difference appears to be less, which is strange.

Detachment of Tightly-bound Electrons from Atoms

We will now consider the most direct and striking evidence for the statement that each atom (apart from those of the lightest elements) possesses several distinct ionizing energies—several distinct "states of ionization." This fact is taken to mean that each atom possesses several or many electrons which are *bound*, as the phrase is, with different degrees of firmness or tightness; that the ionizing-energies of the atom are, so to speak, the *extraction-energies* of these various electrons; to each electron there corresponds a certain extraction-energy, the amount of energy which must be imparted to the atom to extract that electron, the energy-difference between the normal state of the atom and that particular "state of ionization" which involves the absence of that particular electron. I shall frequently use the language of this interpretation, which is extremely convenient and likely to remain so. Nevertheless it is desirable to remember that the quantities actually observed are energy-differences between various states of the atom, or energy-values of various states of the atom referred to the energy-value of the normal state as zero. These energy-values are the data of experience; most other assertions about the states of ionization are speculative.[15]

Conceive a layer of atoms of an element possessing several different values of ionizing-energy W_1, W_2, W_3 and so forth; in other words, atoms which are capable of several states of ionization of which the energy-values exceed that of the normal state by W_1, W_2, W_3 and so forth. Suppose that a beam of radiation of frequency ν, so chosen that the product $h\nu$ exceeds all of the ionizing-energies, falls upon the layer. Such a beam is absorbed as though it consisted of individual particles of energy $h\nu$, each of which is either completely absorbed or totally ignored by the layer of matter upon which it falls. Consider an atom which ab-

[15] In some cases, although not in any which will be discussed in this section, it is found necessary to suppose that several distinct states of ionization correspond to the absence of a particular electron, which is somewhat of a strain upon the picture.

sorbs the amount $h\nu$ of energy from the beam. Through this absorption, an electron is detached from the atom. If however the electrons were merely separated from the atom and left stationary beside it, the energy of the system (ion plus electron) would by definition have been augmented merely by W_i. This quantity is (by our supposition) less than $h\nu$. However the entire energy $h\nu$ has been absorbed; the difference $(h\nu - W_i)$ is likewise transferred to the ion-plus-electron system, in the form of kinetic energy of the liberated electron. The electron flies away with speed V_i determined by the relation

$$\tfrac{1}{2}m V_i^2 = h\nu - W_i. \tag{7}$$

The foregoing paragraph contains several interlocking assumptions, which if they are all true lead to this conclusion: *When a beam of radiation of frequency ν falls upon a layer of atoms having ionizing-energies W_1, W_2, . . . W_i . . . , electrons of various speeds spring out of the layer,*

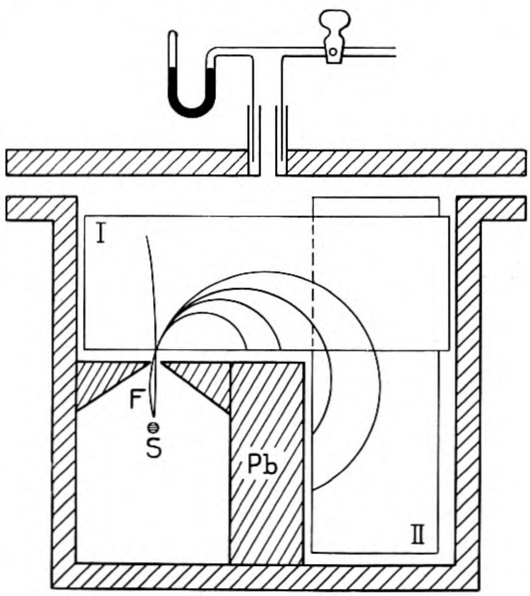


Fig. 6

there being for each value of W, a corresponding group of electrons of which the speed is given in terms of W by equation (7).

Suppose that one irradiates a metal with high-frequency radiation, and by a system of slits confines his experimentation to electrons projected in directions nearly normal to the metal surface, and applies a magnetic field in a direction parallel to the surface. Then we have the situation which occurs in measuring the speeds and charge-to-mass ratios of electrons and ions by the method of electric acceleration followed by magnetic deflection. The only differences are, that in the present case the speeds v with which the electrons enter into the magnetic field are imparted to them not by an imposed electric field but by

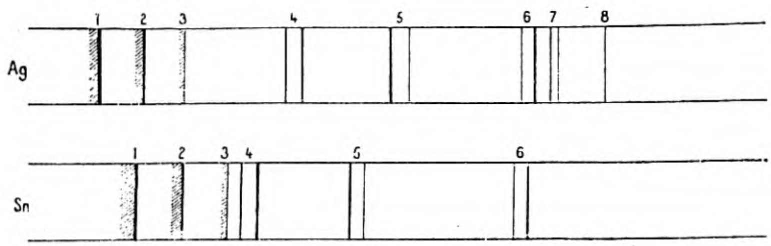


Fig. 7

the radiation which released them; and that the experimenter takes the value of e/m for granted and computes the values of v from the magnetic deflections alone. The electrons are swept around in circular arcs, of which the radii yield their speeds.

The apparatus by which such experiments are performed is of the type shown in Fig. 6. At S there is a long narrow rod or tube of the

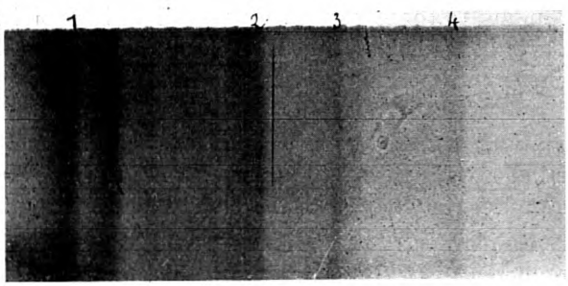


Fig. 8

material to be tested; it is irradiated by X-rays proceeding from a source beyond the diagram to the left. A magnetic field, directed normally to the plane of the paper, sweeps the emerging electrons around in circular arcs, some of which pass through the slit. The appearance of films laid along the top of the block Pb, normal to the plane of the paper, is shown by Figs. 8 and 9. They suggest spectra; and though the lines are signatures of special electron-speeds rather than of special radiation-frequencies, the difference between these is not so radical as once it seemed, and we may without hesitation call them by some such name as *electronic spectra*.

Each line in such a spectrum is produced by electrons of a definite extraction-energy, extracted by radiation of a definite frequency. Continuing with the policy of referring to electrons with a definite extraction-energy as being definitely individualized within the atom, I will designate the electrons of greatest extraction-energy for any particular kind of atom as the K electrons; those of next greatest extraction-energy as the L electrons, and then the M and N electrons in due order. (Later it will be necessary to subdivide these classes, but for the moment this may be avoided.) At other times I shall speak of these elec-

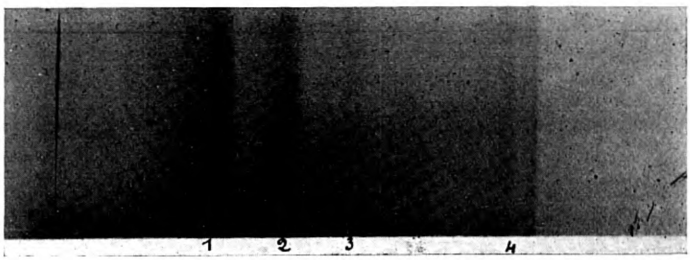


Fig. 9

trons as belonging to the K level, the L level, and so forth; still other terms in use are the K shell and the L shell, or the K ring and the L ring. In a given electronic spectrum we may expect to find a set of lines due to K, L, M and other electrons, for each frequency represented in the incident radiation; unless there are some of these frequencies for which the quantum energy $h\nu$ is less than the extraction-energies of some of the electron-groups, in which case there will be no corresponding lines.

An ideally simple electronic spectrum would be produced by a single radiation-frequency; but this is impracticable, for even if one were to eliminate from the stream of X-rays proceeding out of an X-ray tube all but one frequency, the irradiated atoms would themselves supply others.[16] As in the mass-spectra upon Aston's plates, this unavoidable complexity is actually an advantage; it helps in identifying the several lines.

In Figs. 8 and 9, photographs taken in the manner already mentioned, there appears the electronic spectrum due to silver atoms irra-

[16] These are in fact especially efficient in ejecting electrons, as they originate within the atom-layer itself.

diated by the characteristic X-rays of tungsten.[17] To guard against the possibility that the photographs may lose in clearness by the process of reproduction, I will base the explanation upon the uppermost of the sketches in Fig. 7, which is abstracted by de Broglie from similar pictures. The electron-speeds corresponding to the lines increase from left to right. The irradiating X-rays consist of four characteristic frequencies from the X-ray spectrum of tungsten; in order of decreasing frequency they are known as $K\gamma$, $K\beta$, and the two members of the $K\alpha$ doublet. The four lines marked 4 and 5 in the electronic spectrum are made by electrons extracted by these four radiations from a single level —the K level of the silver atoms. The two following doublets, marked 6 and 7, are made by electrons extracted by the $K\alpha$ frequencies from two other levels of the silver atom, the L and M levels respectively. Line 8 is due to $K\beta$ extracting electrons from the L level. At the other end of the spectrum, the three lines 1, 2, 3 are due to electrons ejected from the L and the M levels by two of the X-ray frequencies characteristic of silver, which the irradiating X-rays stimulate some of the silver atoms to emit. The rays responsible for these particular lines are the so-called $K\alpha$ and $K\beta$ rays of silver, which are so related to one another (as will be stressed in a later passage) that the electrons extracted by the former from the M level have very nearly the same energy as the electrons extracted by the latter from the L level, so that the two frequencies acting on the two groups of electrons produce three (instead of four) distinct lines of the electronic spectrum.

Reverting now to the photographs: in Fig. 8 the pairs of lines marked 4, 3, and 2 are those designated respectively as 6, 5 and 4 in the sketch and in the foregoing explanation, while the lines to the left are those produced by characteristic X-rays of silver acting upon silver atoms. On a larger scale, this latter region of the spectrum is shown in Fig. 9; here the lines are marked by the same numerals as in the sketch; the pair at 4 is due to K-electrons extracted by the two $K\alpha$ rays of tungsten, the line 3 is due to M-electrons extracted by the $K\beta$ radiation of silver, the line 2 results jointly from L-electrons extracted by the $K\beta$ radiation of silver and M-electrons expelled by the $K\alpha$-radiation of silver, while the line 1 is due to L-electrons ejected by the $K\alpha$-rays of silver.

The resemblance and the differences between electronic spectra of elements not far apart in the Procession are illustrated by the two sketches in Fig. 7, the lower relating to tin (atomic number 50) and the upper to silver (atomic number 47) irradiated by the same frequencies.

[17] I am greatly indebted to M. de Broglie for sending me the negatives of these admirable pictures, as well as that of Fig. 10.

Since the extraction-energy of each named class of electrons increases along the periodic table, the lines designated as 4, 5 and 6 in the electronic spectrum of silver reappear in that of tin, displaced in the direction of diminishing electron-speeds, that is, to the left. But, as to the lines 1, 2, and 3, both the extraction-energies of the electrons and the frequencies of the rays responsible for these alter as one passes from silver to tin, and the net result of the double alteration is that the lines are displaced to the right.

The energy-values of the various states of ionization of an atom—or, in terms of the customary picture, the extraction-energies of the various classes of electrons within the atom,—may be determined with a certain degree of precision from experiments such as these. However, as in the case of the measurement of charge-to-mass ratios for individual ions by the methods of Aston and Dempster, there is little incentive to develop the accuracy of the method to the highest possible extent; for most of the energy-values in question can be determined with very great accuracy in another way, which we will now examine.

Absorption of Radiation Through Ionization

When a beam of radiation of frequency ν is transmitted through a layer of matter, from the atoms of which it extracts electrons with an expenditure of energy $h\nu$ at each extraction, we should expect to find it correspondingly reduced in intensity when it emerges from the layer.

This effect is strikingly conspicuous with radiation high enough in frequency to detach the tightly-bound electrons of massive atoms. Let a narrow beam of "heterogeneous" radiation, containing all frequencies throughout the widest possible range, fall from an X-ray tube through slits and diaphragms upon a thin layer of such atoms; let the transmitted rays be dispersed by some appropriate spectroscope, and fall finally upon a photographic plate on which their spectrum—in the ordinary sense of the word, not in the sense of "electronic spectrum"—is outspread.

In Fig. 10 there are three such spectra, of heterogeneous beams which have passed through layers of cadmium, antimony, and barium respectively. The frequency increases from right to left. The darkening at any point is a measure of the intensity with which the X-rays acted at that point.

Below a certain frequency identical for all three elements, the photographic films have evidently been little affected; as soon as this critical frequency is exceeded, the effect suddenly becomes enormous. This critical frequency is the one for which the quantum-energy just suffices to extract a K-electron from a silver atom; for the photographic film contains silver, and it is the expulsion of electrons from the atoms in it which initiates the photographic process. Proceeding always toward higher frequencies, we see that presently the plates suddenly become whiter, at another critical frequency which however

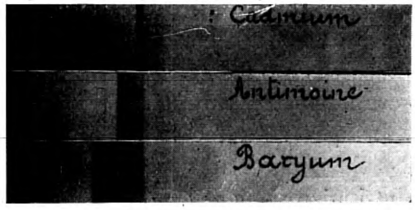


Fig. 10

is not the same for the three elements. The photographic film is not responsible for these "absorption-edges" as they are called; each of them occurs at the particular frequency for which the quantum-energy just suffices to extract a K-electron from an atom of the element which formed the absorbing-layer placed in the path of the beam before it reached the plate. To the right of the absorption-edge we have the lower frequencies, unimpeded by the cadmium (or antimony, or barium) atoms because unable to ionize them; to the left we have the higher frequencies, reduced in intensity by the intercalated matter because some of their energy was drawn off to detach electrons.

From the frequency ν at such an absorption-edge, the extraction-energy W of the class of electrons in question for the kind of atom in question is determined by the equation

$$h\nu = W.$$

This is a much more delicate way of measuring extraction-energies than the observations upon electronic spectra afford. Nevertheless the measurements upon the energies of the ejected electrons are of the greatest importance, for they show what is effected by the energy-transformations which set in when one or another of these critical frequencies is overpassed.

Likelihood of Ionization by Electrons Having More Than the Least Ionizing-Energy

We have seen that electrons projected into a gas of ionizing-energy V_o are able to ionize it if their kinetic energy exceeds V_o, otherwise not (apart from ionizations effected upon atoms in abnormal states). This question now suggests itself: Suppose that a great number Q of electrons, all having kinetic energy V, falls upon a thin stratum of gas containing dN atoms per unit area: how many atoms will they ionize, how many ions will be produced? Designating this number by $Qf(V)dN$: what is $f(V)$?

This question is much easier to formulate in words than to answer by experiment. Suppose for instance that one should try to answer it by means of the scheme of apparatus sketched in Fig. 1. In going from G_1 to G_2, coming to a stop between G_2 and G_3, and returning again, the electrons pass successively through all values of kinetic energy from their highest down to zero and back to their highest again; and ions are produced by electrons of all values of kinetic energy, from their highest down to the ionizing-energy. The ions collected by the collector at C represent a sort of integral of $Qf(V)dN$ taken between V_o as minimum and the energy possessed by the electrons at G_1 as maximum. To determine $Qf(V)dN$ it is necessary to measure the total ionization at several values of V and then construct a sort of differential curve. To determine Q it is necessary to know how many electrons come from the filament into the ionizing-region, and in addition how many extra ones are introduced through primary electrons knocking them out of the gauze of G_1.

Another scheme consists essentially in making G_2 into a solid wall and using it to collect the electrons, so that after passing from G_1 to G_2 they vanish from the scene. This would be excellent if the region between G_1 and G_2 could be left equipotential; but it is necessary to intrude a negatively-charged electrode in order to collect the positive ions, and apparently whenever this electrode is sufficiently large and sufficiently negative to capture the ions it is also sufficiently large and sufficiently negative to distort the field between G_1 and G_2 quite seriously; so that the electrons are at first slowed down and later speeded up again as they pass from G_1 to G_2, and the ions received by the collector are as before a sort of integral of $Qf(V)dN$.

In spite of these difficulties the various experiments performed with extremely rarefied gases yield fairly concordant results.[18] The function $f(V)$ mounts steadily, from zero at the ionizing-energy V_o, to a broad and flattish peak culminating somewhere between 100 and 400 volts (depending on the gas), and thereafter declines slowly as V increases. Thus, although an electron striking an atom (or molecule) can detach the loosest electron if it has just the requisite energy, its chance of doing so is improved if its energy is greater than the just-sufficient

[18] K. T. Compton and C. C. van Voorhis, *Phys. Rev.* (2) 26, pp. 436–453 (1925) and literature there cited; also W. P. Jesse, ibid. pp. 208–220.

amount. However, it would not be safe to infer that throughout the range of these observations all of the ionizations consist in detachments of valence-electrons from various atoms. Sooner or later transfers of atoms into other states of ionization must commence. This is rendered all the more probable by the fact that the values of $f(V)$, determined at or near the peak for each gas, show a very definite tendency to increase steadily with the number of electrons in the atom or the molecule in question.

If a stream of electrons is projected into a sufficiently dense gas, the electrons are gradually slowed down and even stopped, and the stream is dispersed. Measurements of the number of ions produced per electron per millimetre have been made under such conditions,and measurements also of the "total ionization" produced in a volume of gas so large that the electrons lose their forward speed altogether before reaching the walls; but though the intrinsic interest of such measurements is great, it seems practically impossible to deduce $f(V)$ from them.[19] The difficulties may be compared with those arising in the study of alpha-particle scattering when the metal foil is too thick.[20] When, however, the electrons are moving with the enormous speeds possessed by those ejected from radio-active substances, or when ionization by alpha-particles is studied, the conditions again become simpler and relatively intelligible.

Ionization by Alpha-Particles and Very Fast Electrons

Ionization by particles possessing kinetic energies amounting to millions of equivalent volts, such as alpha-particles and many of the electrons emerging from radioactive substances, might well be expected to follow other laws than ionization by particles possessing little more than enough energy to detach an electron from an atom. Such indeed is the case; yet it would not be justified, either by reasoning or by experiment, to suppose that even such highly energetic particles expel electrons of any and every class tightly-bound alike and loosely-bound alike, with equal ease and abundance from the atoms which they strike.

It is not particularly difficult to measure the total number of ions produced by an alpha-particle in its course through a gas from the moment it enters, with a measurable initial speed, to the moment when it goes into retirement (so to speak) as an ordinary helium atom; nor to

[19] G. A. Anslow, *Phys. Rev.* (2) 25, pp. 484–500 (1925) and literature there cited.

[20] Anyone desiring to learn how complicated the circumstances may become when electrons are shot into a dense gas should read P. Lenard's brochure "Quantitatives über Kathodenstrahlen," published by the Heidelberg Academy in 1918.

divide this number into the kinetic energy which it originally had, thus obtaining the average energy spent per ion (or rather per pair of ions generated, since each ionization produced two ions of opposite sign)—a quantity amounting generally to several tens of equivalent volts (33 volts for air).[21] By performing such experiments with alpha-particles of various initial speeds, it is possible to determine a function analogous to the function $f(V)$ defined for electrons in a previous section. This function increases rapidly as the aplha-particle approaches the end of its sharply-terminated trail, varying approximately as the reciprocal of the cube root of the distance it has yet to go.

Alpha-particles as they pass through a gas thus produce a countable number of ions and suffer a measurable loss in kinetic energy. It is interesting to enquire whether these two processes can be identified with one another and explained by the nuclear atom-model—whether the lost kinetic energy is altogether spent in detaching electrons from the atoms of the gas and supplying them with extra kinetic energy.

Before comparing any theory with the experimental data, one must be aware of two complexities. In the first place, an alpha-particle may transfer energy to an atom without ionizing it; so that the energy it loses in passing through a gas may exceed that which it spends in ionizing. In the second place, some of the ions produced by an alpha-particle—notably, the detached electrons—may themselves be endowed with energy enough to ionize, so that a measurement of the total ionization in the gas may yield an excessive estimate of the number of ions actually and immediately produced by particles striking atoms. Naturally the energy for producing all of these ions, "primary" and "secondary" alike, comes from the alpha-particles, so that such data as the aforesaid values for energy-spent-per-ion-generated have a definite meaning.

Discrimination between ions produced directly and indirectly is desirable, indeed essential, for testing any theory; but thus far there is no way for distinguishing the two, except in the case of very fast electrons for which the trails have been photographed with great magnification by C. T. R. Wilson[22] by his celebrated expansion-method, in which each ion formed in the passage of such an electron through a gas becomes the center of a visible droplet of water. In some of his pictures, in Fig. 11, pairs of droplets and also groups of four, six and more are seen. The paired droplets have condensed upon the two ions, positive and negative, produced by a single primary ionization (and

[21] R. W. Gurney, *Proc. Roy. Soc.* A107, pp. 332–340 (1925) and literature there cited.

[22] *Proc. Roy. Soc.* A104, pp. 192–212 (1923).

then drawn apart by an appropriate electric field); the groups of more than two bear witness of a primary ionization followed by secondary processes of the same type. In Fig. 12 there is an actual long branch to the primary trail; the original fast electron has detached another and endowed it with so great an energy that in ionizing-efficiency it

Fig. 11

rivals its liberator. These figures show that a mere count of all the ions formed by a particle flying through a gas is no estimate of the detachments of electrons from atoms which the particle of itself and at first hand effected.

Various theoretical expressions have been derived for the rate of slowing-down and the rate of ionization of an alpha-particle or fast

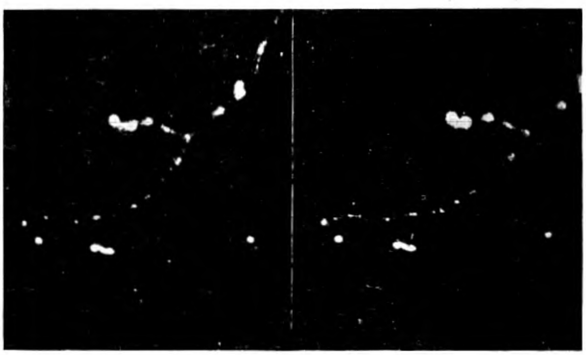


Fig. 12

electron proceeding through a gas. Most of them lead to what are known as "order-of-magnitude agreements," but none to a close quantitative agreement—which is, perhaps, after all better than could be expected. They are founded upon an equation originally proposed by J. J. Thomson. Suppose a stratum of an element of atomic number Z, containing N atoms; using the nuclear atom-model, we conceive this as a region containing N nuclei and NZ electrons. If the electrons (of mass m) were free and stationary, an alpha-particle of mass M moving with speed U along a line passing at distance p from the initial position of any one of them would communicate to it an amount of energy:

$$W=\frac{8e^2}{mU^2(p^2+a^2)}$$

where

$$a=\frac{2e^2(M+m)}{mUM^2}. \tag{8}$$

Imagine Q alpha-particles passing through this collection of NZ electrons; the number of encounters for which this energy-value lies between two values W and $W+dW$ is equal to $2\pi p(dp/dW)dW$. Multiplying this by W and integrating over all values from $W=0$ (corresponding to $p=\infty$) to $W=8e^2/mU^2a^2$ (corresponding to $p=0$), we arrive at a value for the total amount of energy communicated by the alpha-particles to the electrons, which value is infinite. This absurd conclusion rests on the absurd assumption that the electrons are free, which, of course, is not made. Generally it is assumed that whenever p exceeds a certain value, selected for one reason or another, equation (8) loses its validity and W is zero; for instance, that whenever p is so great that the value computed by (8) for W is smaller than the least energy sufficing to remove the electron altogether from the atom or to put the atom into a Stationary State, then there is no transfer of energy whatever; but, whenever p is so small that W as computed by (8) exceeds the extraction-energy for the electron in question, then the electron is extracted and carries off, as kinetic energy, the difference between W and its extraction-energy.

Definite assumptions must be made about the extraction-energies of the various classes of electrons in the atom, the number of electrons in each class, and the Stationary States of the atom; this being done, formulae are derived for the primary ionization, the scondary ionization, and the rate at which the alpha-particle (or fast electron) loses energy.[23] Apart from these results of elaborate and careful analysis which lead as I have said to order-of-magnitude agreements (in some cases the agreements approach quantitative value) it may be pointed out that the equation (8) leads, when U is so great that a becomes small relatively to p, to the conclusion that as a fast-flying particle proceeds through matter the fourth power of its speed falls off linearly with increase of distance traversed, which is in agreement with much

[23] See R. H. Fowler, *Proc. Camb. Phil. Soc.* 21, pp. 521–540 (1923), and G. H. Henderson, *Phil. Mag.* 44, pp. 680–(1922) for discussion and prior literature as well for their own work.

experimental work. It furthermore indicates that the total ionization effected by a beam of particles in traversing a given thickness of matter should, beyond a certain speed, diminish with increasing speed; which for alpha-particles is true for the entire available speed-range, and for electrons is true beyond the speed of optimum ionizing-efficiency mentioned in a previous section.[24]

Multiple Ionization

The analyses of positive rays issuing from gases sustaining electrical discharges show that under such conditions some atoms are deprived of two, three, or even so many as eight electrons. The recently-developed methods of interpreting spectra make it practically certain that some of the spectrum-lines emitted from gases bombarded by electrons or sustaining discharges, and particularly from the exceptionally violent discharges known as "sparks," are due to atoms lacking one, two, or so many as six of their normal complement of electrons.[25]

Such ions might conceivably be produced either in one operation or in several; that is, the two (or more) missing electrons might have been removed by a single agency at a single moment, or they might have been detached one after the other by separately and successively acting agents. Measurements by the method of H. D. Smyth, such as those upon argon already cited, are capable of showing the minimum amount of energy which bombarding electrons must possess, in order that doubly-ionized atoms may appear in a bombarded gas; but they do not show, at least not directly, whether this minimum amount is what is required to effect double ionization in a single operation, or merely what is required to effect the most difficult among two or several steps leading cumulatively to the result. The same holds true about the experiments in which the least bombarding-voltage sufficient to bring out the spectrum associated with the doubly-ionized atom is measured.[26] Granted that the energy-difference between the once-ionized and the normal argon atom is 15 equivalent volts, and

[24] The attempts to account for "straggling" of alpha-particles—that is, for the fact that different particles of the same initial speed are slowed down at somewhat different rates in progressing through the same gas—by ascribing it to mere statistical fluctuations in the number of electrons close to which they passed seem to have been unsuccessful; the observed straggling is much too great for this explanation. See G. H. Henderson, *Phil. Mag.* 44.

[25] Multiply-ionized atoms are regularly observed in electrolytic solutions of compounds of other-than-monovalent elements; strangely enough they are rarely if ever found among the ions issuing spontaneously from hot metals and salts.

[26] P. D. Foote et al., *Phil. Mag.* 42, pp. 1002–1015 (1921); *Astroph. Jour.* 55, pp. 145–161 (1922); *Origin of Spectra*, 1922.

that A^{++} ions appear in argon bombarded by 45-volt electrons: do 45 equivalent volts constitute the amount of energy necessary to remove two electrons at once from a normal atom, or the amount necessary to remove one electron from an atom which a prior electron-impact has ionized? The question is not different in principle from one arising in measurements of the first ionizing-potential, whether the first appearance of ions signifies simply that atoms are being ionized in two stages; but apparently it is harder to settle by direct evidence.

Analysis of the spectra of the ion and of the atom whenever practicable, discloses definitely the energy-differences between the state of double ionization, the state of single ionization, and the normal state of the neutral atom. Thus with helium the first of these is greater than the second by 54 and then the third by 79 equivalent volts. Similar calculations for magnesium show that the first is greater than the second by 15 equivalent volts; as the spectrum of the ion Mg^+ in Foote's just-cited experiments appeared at about that energy of the bombarding electrons, the atoms in the vapor must have been ionized by two successive impacts.

In the course of R. A. Millikan's observations upon droplets of oil floating in ionized gases, he found that they never captured charges amounting to $2e$ or a greater multiple of e, except in the solitary instance of helium traversed by alpha-particles; in this case about one out of every six positive charges captured was a double electron-charge $2e$. He concluded that not more than one electron was ever detached from an atom in a single operation, except that among encounters of alpha-particles with helium atoms about one-sixth caused both of the electrons of the struck atom to be torn away.[27]

Detachments of two or more tightly-bound electrons from a massive atom, whether effected in one operation or in several, might be revealed by additional absorption-edges in the spectrum of an X-ray beam after passing through matter; certain delicate features in X-ray spectra have in fact been explained in this manner.

Thermal Ionization [28]

In addition to all the information about ionization by particular agents such as electrons of specified speeds and radiation of specified frequencies, there is reason for making certain assertions about ioniza-

[27] The percentage may well have been much greater, since many of the ions left behind after the passage of the alpha-particle were probably produced by secondary, not primary ionization (R. H. Fowler).

[28] General references: E. A. Milne, *Proc. Phys. Soc. London* 36, pp. 94–113, and literature there cited; A. A. Noyes, H. A. Wilson, *Pro. Nat. Acad. Sci.* 8, pp. 303–307 (1922).

tion *per se*, apart from all knowledge or assumption concerning the processes which effect it. There is a thermodynamic method of determining the percentage of dissociated molecules in a molecular gas as a function of the temperature and the pressure of the gas, which can be used if we know the amount of energy required to dissociate a single molecule, the specific heats of the undissociated and the dissociated gas, and the chemical constants of the undissociated and the dissociated gas. An analogy may be established between dissociation and ionization: the ionizing-energy of a monatomic gas corresponds to the heat of dissociation of (say) a diatomic gas; the electrons and the ions resulting from the ionizations may be taken as the particles of two distinct gases mingled with one another and with the gas composed of the neutral atoms; the chemical constant of the ion-gas is taken as equal to the chemical constant of the original gas, and the chemical constant of the electron-gas is identified with that which a gas composed of neutral atoms, each possessing the same mass as an electron, would possess. Utilizing this analogy, a formula may be deduced for the percentage of ionized atoms present in a monatomic gas in thermal equilibrium at any temperature and pressure.

Without developing the formula, it may be taken as a rather obvious inference that the higher the temperature of the gas at a given pressure, or the lower the pressure at a given temperature, the greater the percentage of ionization will be; and of two gases maintained at the same temperature and pressure, the gas having the smaller ionizing-energy will be the more ionized.

Measurements of the degree of ionization in a flame of known temperature, into which a known amount of caesium was introduced, have yielded values in good agreement with the percentage calculated from the thermodynamic formula; and measurements upon the conductivities of the vapors of the alkali metals have shown that they stand in the order of the ionizing energies reversed, although in other respects the agreement with the theory is not good.[29]. The tests and the value of the theory, however, appear chiefly in the realm of astrophysics. The hotter the region of a star in which the lines observed in its spectrum have their source, the more the lines of ionized atoms predominate among these. In many cases it happens that lines of ionized atoms are the only ones characteristic of a given element to be found at all. The assertion once commonly made, that certain elements are absent from the sun or other stars, is invalidated by the fact that under the actual conditions of temperature and pressure prevailing in these bodies,

[29] B. T. Barnes, Phys. Rev. (2) 23, pp. 178–188 (1924); M. N. Saha, Phil. Mag. 46, pp. 534–543 (1923).

those elements if present would be totally ionized and would not reveal their familiar lines at all. Rubidium was thought to be omitted from the composition of the sun, until it occurred to H. N. Russell to look for its lines in the spectra of comparatively cool sunspots. The relative intensities of the lines of ionized and non-ionized atoms of various kinds in the spectra of individual stars are now ascertained and used as a guide in assigning temperatures to these stars, and their guidance is shown reliable by the accord between the conclusions to which it leads and conclusions otherwise attained. The study of ionization in the laboratory thus contributes to the understanding of the stars.

Methods of High Quality Recording and Reproducing of Music and Speech Based on Telephone Research[1]

By J. P. MAXFIELD and H. C. HARRISON

SYNOPSIS: This paper deals with an analysis of the general requirements of recording and reproducing sound without appreciable distortion. The storing or recording of sound requires, first, a mechanical system which will respond faithfully to the sound waves which are to be recorded. Then there is required some material in or on which this sound may be recorded and an intervening system which permits the sound waves to make the record in this material. In the usual case, and in that which is particularly discussed, there is a mechanical system which will vibrate in response to the sound which is to be recorded and directly through some mechanical linkage, or less directly through an electrical linkage, drives a cutting mechanism which will impress a wax record.

The amount of power available to operate the recorder directly from the sound in the recording room is so small as to make the use of high quality electrical apparatus with associated vacuum tube amplifiers of very distinct advantage over the acoustic method.

Where the question of reproduction is concerned, the same two alternatives mentioned for recording present themselves, namely, direct use of power derived from the record itself vs. the use of electro-mechanical equipment with an amplifier. In this case, however, the situation is materially different since the power which can be drawn directly from the record is more than sufficient for many uses. It is, therefore, generally simpler to design one single mechanical transmission system than it is to add the unnecessary complications of amplifiers, power supply and associated circuits. In cases where music is to be reproduced in large auditoriums, the power which can be drawn from the record may be insufficient and some form of electrical reproduction using amplifiers becomes necessary.

The paper points out, at length, how many of the heretofore unsolved fundamental problems of sound recording and reproduction have been readily solved by the application of a detailed knowledge of telephone transmission theory. The advances which have been effected in telephone transmission theory and in related electrical measuring apparatus in the last few years, have been so great as to surpass previous knowledge of mechanical wave transmission systems. The result is, therefore, that mechanical transmission systems of the type here considered, and perhaps other types, can be designed more successfully if they are viewed as the analogs of electric circuits. A detailed analysis is here made of the analogies between electrical and mechanical systems in the voice frequency range and a discussion of the resulting mechanical design is presented.

INTRODUCTION

THE problem with which this paper is concerned, in its broadest sense, may be stated as that of taking sound from the air, storing it in some permanent way and reproducing it again without appreciable distortion. It is immaterial from the general standpoint whether the means used are mechanical or electrical or a combination of the two. The choice of which method to use will depend largely upon the commercial requirements accompanying the specific purpose for which the reproduction is being made. For instance, it is quite probable that

[1] As printed here this paper is essentially as read before the A.I.E.E. Feb. 8-11, 1926.

the means chosen for reproduction in residences would differ materially from those used in large ballrooms or in the presentation of synchronized motion pictures.

Before considering the methods and results referred to in the title of this paper, it may be well to make a rough division of the problem. The storing or recording of sound requires, first, a mechanical system which will respond faithfully to the sound waves which are to be recorded. Then, there is required some material in or on which this sound may be recorded and an intervening system which permits the sound waves to make the record in this material. In the usual case, and in that with which we are particularly concerned here, there is a mechanical system which will vibrate in response to the sound which is to be recorded and directly through some mechanical linkage or less directly through an electrical linkage, drive a cutting mechanism which will impress a wax record.

The first consideration, therefore, is the character of the sound which is to be recorded including all of the effects of reverberation and the general questions of studio design. Next to be considered is the manner in which the cutting instrument shall impress this speech or musical record upon the constantly rotating wax disk, which disk is commonly called the wax master. In this connection, there will be discussed also the relative value of the electrical and mechanical linking of the cutting knife with the mechanism which receives the sound waves. Following the discussion of these problems and a brief reference to the state of the prior art, there remains to be considered the reproduction of the sound which is stored in the cuts or grooves of the wax record.

In the case of reproduction also, there is required a mechanical system which will respond to these cuts in the wax and a system which will set up in the air-sound waves essentially identical to those picked up by the first mechanism of the recording system. Between these two systems, a mechanical linkage intervenes in the case under discussion, but reference is made to the relative advantages of this system compared with the use of an electrical linkage.

First to be described, is the character of the sound which is to be recorded and reproduced and the effects of reverberation and transients upon the listener's sensation of this sound.

Studio Characteristics and Transients

Phonographic reproduction may be termed perfect when the components of the reproduced sound reaching the ears of the actual listener have the same relative intensity and phase relation as the sound reaching the ears of an imaginary listener to the original performance would have had. Obviously, it is very difficult, if not impossible, to fulfill all of these requirements with a single channel system, that is, with a system which does not have a separate path to each ear of the listener from the sound source.

The use of two ears, that is, two-channel listening, gives the listener a sense of direction for each of the various sources of sound to which at a given moment he may be listening, and, therefore, he apprehends them in their relative distribution in space. It has been found possible with a single channel system, however, by controlling the acoustic properties of the room in which the sound is being recorded, to simulate to a considerable degree in the reproduced music the effective space relationships of the original. In this case, with a one-channel system, the directional effect is, of course, entirely absent, and the spatial relationship which is apprehended is probably due to the increased apparent reverberation of the instruments situated at the far end of the room as compared with those in the near foreground.

In recording work, therefore, one of the important acoustic characteristics of a room is its time of reverberation. Although it is probable that this is the most comprehensive single factor, experiment has shown that the shape of the room and the distribution and character of the damping surfaces play a part in the excellence of music in such a room.

It has been shown by Sabine[2] that for piano music, studios should have a time of reverberation measured by his method of 1.08 seconds. Experience has indicated that this figure is also very closely correct for other types of music. This figure of Sabine's assumes binaural listening. With single-channel systems, such as most of the present reproduction systems, whether for radio or the phonograph, the ability of the listener to separate the reverberation from the direct music by means of the sense of direction is completely removed and there is thrust upon his attention an apparently excessive amount of room echo. Experiment has shown that a time of reverberation for the recording room ranging from slightly more than ½ to slightly less than ¾ of Sabine's figure affords in the reproduced music the effect of a room with proper acoustics. When this effect is accomplished, the person listening to the reproduced music has the consciousness of the music being played in a continuation of the same room in which he is listening and also has a sense of spatial depth.

Experiment has indicated further that any transients set up by the recording or reproducing system constitute a second cause of apparent

[2] Collected papers of W. Sabine.

increased reverberation. The data obtained thus far are insufficient to permit assignment of quantitative values to the importance of these two factors.

At the present state of the art, the most important requirement of a recording or reproducing system is its frequency characteristic. This involves two factors—intensity versus frequency, and phase distortion versus frequency. The effect of the second of these factors is not thoroughly understood but as it is closely related to the production of transients it has to be considered, as mentoined above. The system to be described is, however, relatively free from violent phase shifts within most of the range covered, but does have some undesirable phase-shift characteristics with small accompanying transients near its limiting cut-off frequencies.

Frequency Requirements

The frequency range which it would be desirable to cover if, it were possible, with relatively uniform intensity for the transmission of speech and all types of music including pipe organ is from about 16 cycles per second to approximately 10,000.

It may be interesting to examine the record requirements for a band of frequencies this great. For the purpose of this illustration, a lateral cut record will be assumed although in all the factors except the time which the record will run, the arguments apply in a similar manner to the hill-and-dale cut. Since, for mechanical reproduction, the sound at a given pitch is radiated by means of a fixed radiation resistance, it is necessary that the record must be cut with a device the square of whose velocity is proportional to the sound power. Under these conditions, it is seen that for a given intensity of sound the amplitude is inversely proportional to the frequency of the tone, and that a point will be reached somewhere at the low end of the sound spectrum where this amplitude will be great enough to cut from one groove into the adjacent groove, or in case of vertical cut, to cut so deeply that with present materials the wax will tear instead of cut away with a clean surface. This means that there is an inherent maximum amplitude beyond which it is not commercially feasible to go. Similarly the minimum radius of curvature of sine waves of various frequencies cut at constant velocity is inversely proportional to the frequency, so that as higher and higher frequencies are reached the radius of curvature becomes smaller and smaller until finally it becomes too small for the reproducing needle to follow. There is, therefore, an inherent limit at the upper end.

In order to extend these limits, it is necessary in the case of the low end to make the spiral coarser and in the case of the high end to run the record at a higher speed. Both of these changes tend to decrease the time which a record of a given size can be made to play. The only alternative of these methods is to cut the record less loud than is the present standard practise and make the reproducing equipment more sensitive. This could easily be done if it were not for the "record noise" or "surface noise," as it is commonly called. Since this surface noise is already loud enough in comparison with the reproduced music to be somewhat objectionable, no appreciable gain in this direction can be made until the technique of record manufacture has been distinctly improved.

In this connection, there is one other interesting point. It has been suggested that if electric reproduction were used, it would be possible to cut the record with a characteristic other than uniform velocity sensitiveness and correct for the error by an electrical system whose characteristic is the inverse of the characteristic of record. If the change which is made in the recording characteristic tends toward cutting at uniform acceleration sensitiveness, the amplitude varies inversely as the square of the frequency and hence the difficulties at the low end of the scale are greatly enhanced. Similarly, if the records are cut more nearly at constant amplitude, the radius of curvature of the sine waves decreases as the square of the frequency, hence the difficulties are placed at the upper end. In the process which is being described in this paper, these limitations have been met commercially by having a frequency characteristic of the uniform velocity type between the frequencies of 200 and approximately 4000 cycles per second. Below 200 it has been necessary to operate at approximately constant amplitude with a resulting loss in intensity which loss increases as the frequency decreases. Above 4000 it has been necessary to operate at approximately constant acceleration with its consequent slight loss in intensity at the very high overtones. With a characteristic of this type, a range of frequencies from 60 cycles to 6000 can be recorded with reasonable success although the very low and very high range are slightly deficient. (See Fig. 14) With a record having such a frequency characteristic, the inherent limitations are divided between the two ends of the frequency band and where electrical reproduction methods are used, it is possible to employ a reproduction system whose frequency characteristic compensates for that of the record.

It should be pointed out that an attempt to record notes lower than the low cutoff of the above mentioned apparatus would result in recording only those harmonics of the notes which lie above the cut-off. This in no way prevents the listener from hearing the notes, reproduced by means of the harmonics only, as notes with the pitches of the missing fundamentals although it does somewhat change the quality of the tone.[3] It it were not for this ability of the ear to add the fundamental pitch of a note, of which only the harmonics are being reproduced, most of the older phonographs and loud speakers would have been totally useless for the reproduction of speech and music.

Mechanical Versus Electrical Recording

In attacking the recording part of the problem, two ways at once present themselves; first, the direct use of the power of the sound being recorded to operate the recording instrument; and second, the use of high quality electric apparatus with vacuum tube amplifiers in order to give more freedom to the artists and better control to the process. The amount of power available to operate the recorder directly from the sound in the recording room is so small as to make it extremely difficult to make records under natural conditions of speaking, singing,

Fig. 1a—Picture of an orchestra recording by the acoustic process. This picture was furnished through the courtesy of the Victor Talking Machine Company, Camden, New Jersey

or instrumental playing. As the use of high quality electric apparatus with associated amplifiers has a very distinct advantage over the acoustic method, they have been adopted for the recording part of the process. Fig. 1a shows a picture of a group of artists recording by

[3] Physical Criterion for Determining the Pitch of a Musical Tone, H. Fletcher *Phys. Rev.*, Vol. 23, No. 3, March, 1924.

means of the sound power directly, while Fig. 1b shows a record being made by the same artists with the electric process.

It will be noticed in Fig. 1a that the artists are grouped very closely about the horn. In the case of the weaker instruments such as violins, it has been possible to use only two of standard construction. The rest of the violins are of the type known as the "Stroh" violin which is a device strung in the manner of a violin but so arranged that the bridge

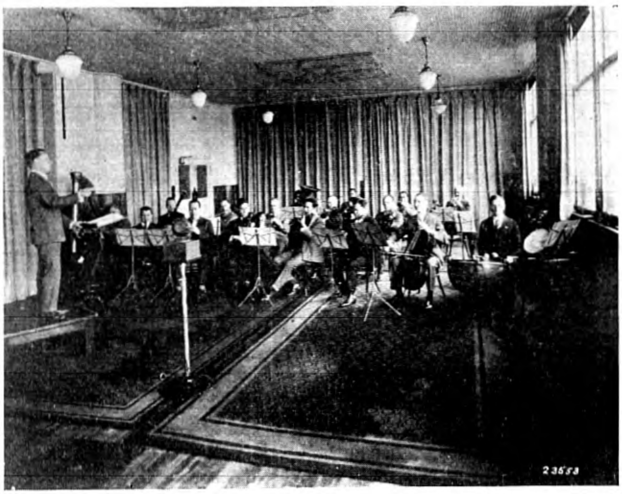


Fig. 1b—Picture of the same orchestra shown in Fig. 1a, but recording by the electric process. This picture was furnished through the courtesy of the Victor Talking Machine Company, Camden, New Jersey

vibrates a diaphragm attached to a horn. The horn is directed toward the recording horn, as shown by the player in the foreground.

With such an arrangement of musicians, it is very difficult to arouse the spontaneous enthusiasm which is necessary for the production of really artistic music. In Fig. 1b the musicians are sitting at ease more nearly in their usual arrangement and all are using the instruments which they would use were they playing at a concert. Furthermore, the microphone is now sufficiently far away from the orchestra to receive the sound in much the manner that the ears of a listener in the audience would receive it. In other words, it picks up the sound after it has been properly blended with the reflections from the walls of the room. It is in this way that the so-called "atmosphere" or "room-tone" has been obtained.

In the old process, it sometimes happened that after the instruments had been arranged in such a manner that the relative loudness of the various parts had been balanced correctly, it was found that the whole selection was either too loud or too weak. This usually meant a complete rearrangement of the players. With the flexibility introduced by the use of electrical apparatus including amplifiers, the control of loudness is obtained by simple manipulation of the amplifier system and is in no way related to the difficulties of the relative loudness of one instrument to another. The only problem for the studio director in this case is to obtain the proper balance among the various musical instruments and artists. The advantages derived from this added ease of control are also made manifest in that it is much easier and less tiresome for the artists and it is usually possible to make more records in a given time.

Mechanical Versus Electrical Reproducing

Where the question of reproduction is concerned, the same two alternatives mentioned for recording present themselves, namely, direct use of power derived from the record itself versus the use of electromechanical equipment with an amplifier. In this case, however, the situation is a little different as the power which can be drawn directly from the record is more than sufficient for home use. Since any method of reproducing from mechanical records by electrical means involves the use of a mechanical device for transforming from mechanical to electrical power and a second such device for transforming from electrical back to mechanical power, that is, sound, it is necessary to use two mechanical systems, one at each end of an electrical system. Where the power which can be supplied by the record, is sufficient to produce the necessary sound intensity, as in the case of home use, it is in general simpler to design one single mechanical transmission system than it is to add the unnecessary complications of amplifiers, power supply and associated circuits. In cases where music is to be reproduced in large auditoriums, the power which can be drawn from the record may be insufficient and some form of electric reproduction using amplifiers becomes necessary.

Brief Description of Recording System

The system used for recording consists of a condenser transmitter, a high quality vacuum tube amplifier and an electromagnetic recorder. Fig. 2 shows the calibration of the condenser transmitter and the associated amplifiers. The condenser transmitter and amplifiers are so designed that the current delivered to the recorder circuit is essentially proportional to the sound pressure at the transmitter diaphragm. The electromagnetic recorder, which will be described later, is designed to work with this type of system. With the exception of this electromagnetic recorder, apparatus of this type has already been described in the literature.[4] In addition to this equipment which might be called the

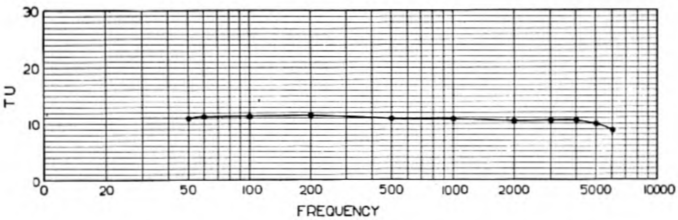


Fig. 2—Calibration of the condenser transmitter and associated amplifiers
This curve shows merely the relative frequency sensitiveness of the system, the zero line having been chosen arbitrarily.

recording amplifier system, there is a volume indicator for measuring the power which is being delivered to the recorder and also an audible monitoring system. The audible monitoring system consists of an amplifier whose input impedance is high compared with the recorder impedance and a suitable loud speaking receiver. The monitoring am-

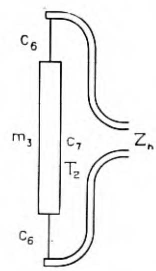


Fig. 3—Schematic mechanical arrangement of diaphragm and air chamber

plifier is bridged directly across the recorder and operates the loud speaking receiver so that the operator may listen to the record as it is being made.

[4] Wente, E. C., "Condenser Transmitter as a Uniformly Sensitive Instrument for Measuring Sound Intensity," *Phys. Rev.*, Vol. 10, 1917.

Crandall, I. B., "Air-Damped Vibrating Systems," *Phys. Rev.*, Vol. 11, 1918.

Wente, E. C., "Electrostatic Transmitter," *Phys. Rev.*, Vol. 19, 1922.

Martin, W. H. and Fletcher H., "High Quality Transmission and Reproduction of Speech and Music," Trans. *A. I. E. E.*, Vol. 43, 1924, p. 384.

Green, I. W. and Maxfield, J. P., "Public Address Systems," Trans. *A. I. E. E.*, Vol. 43, 1923, p. 64.

In the design of the recording and reproducing systems each part of the system has been made as nearly perfect as possible. Errors of one part have not been designed to compensate for inverse errors in another part. Although this method is the more difficult, its flexibility, par-

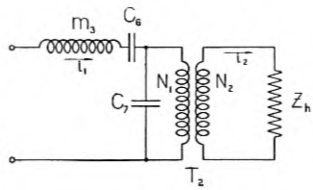


Fig. 4—Electrical equivalent of mechanical system shown in Fig. 3

ticularly as regards the commercial possibilities of future improvements justifies the extra effort.[5] There is, therefore, no distortion in the record whose purpose is to compensate for errors in the reproducing equipment; the only intended distortion in the record being that re-

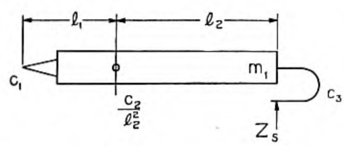


Fig. 5—Schematic mechanical arrangement of needle arm transformer

quired by the inherent limitations mentioned above. See Figs. 2, 14 and 20.

General Basis of Design

An interesting feature of the development of the mechanical and electromechanical portions of the recording and reproducing system is their quantitative design as mechanical analogs of electric circuits. Both the recording and reproducing systems are good examples of the use of this type of analogy.

The economic need for the solution of many of the problems connected with electric wave transmission over long distances coupled with the consequent development of accurate electric measuring apparatus has led to a rather complete theoretical and practical knowledge of electrical wave transmission. The advance has been so great that the knowledge of electric systems has surpassed our previous engineering

[5] Green, I. W. and Maxfield, J. P., "Public Address Systems," Tras. A. I. E. E., Vol. 42, 1923, p. 64.

knowledge of mechanical wave transmission systems. The result is, therefore, that mechanical transmission systems can be designed more successfully if they are viewed as analogs of electric circuits.

While there are mechanical analogs for nearly every form of electrical circuit imaginable, there is one particular class of electrical circuits

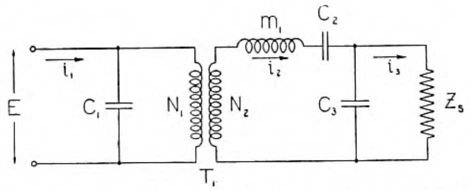


Fig. 6—Electrical equivalent of system shown in Fig. 5 with its termination

whose study has led to ideas of the utmost value in guiding the course of the present development. This class of circuits consists of infinitely repeated similar sections of one or more lumped capacity and inductance elements in series and shunt and are commonly known as filters. The study of filters began with the work of Campbell[6] and a recognition

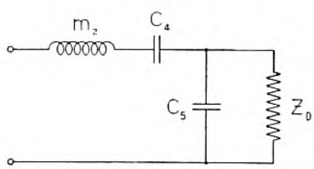


Fig. 7—Electrical equivalent of the spider section

of their importance as frequency selective systems in telephone repeaters, carrier systems, radio, signalling systems, etc., led to their intensive study. In the available literature is to be found a fairly complete statement of their properties and details of their design.[6]

[6] Campbell, G. A., "On Loaded Lines in Telephonic Transmission," *Phil. Mag.*, March 1903.

Campbell, G. A., U. S. Patents 1,227,113; 1,227,114; "Physical Theory of the Electric Wave Filter," *Bell System Technical Journal*, November 1922.

Zobel, O. J.,"Theory and Design of Uniform and Composite Electric Wave Filters," *Bell System Technical Journal*, January 1923.

Peters, L. J., "Theory of Electric Wave Filters Built up of Coupled Circuit Elements," Trans. *A. I. E. E.*, May 1923.

Carson, J. R. and Zobel, O. J., "Transient Oscillations in Electric Wave Filters," *Bell System Technical Journal*, July 1923.

Zobel, O. J., "Transmission Characteristics of Electric Wave Filters," *Bell System Technical Journal*, October 1924.

Johnson, K. S., and Shea, T. E., "Mutual Inductance in Wave Filters with an Introduction on Filter Design," *Bell System Techincal Journal*, January 1925.

Johnson, K. S., "Tansmission Circuits for Telephonic Communication," D. Van Nostrand, 1925.

It will be recalled in the case of the telephone circuit that the introduction of inductance coils at regular intervals in the circuit produced a remarkable change in the transmission characteristic. Over a broad band of frequencies the attenuation was reduced and made fairly uni-

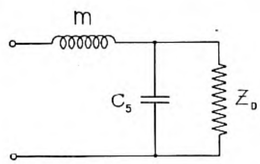


Fig. 8—Electrical equivalent of simple low pass type of network which occurs frequently in this work

form over that range while beyond a critical frequency called the cut-off frequency the attenuation became very high. In the ideal filters with zero dissipation the transmission characteristics are of the same nature but more clear cut. Structures of this type with infinitely repeated sections will have one or more transmission bands of zero attenuation and one or more bands having infinite attenuation. The impedance characteristics of such a structure measured from certain characteristic points will be pure resistance more or less uniform in the transmission bands, and pure reactance in the attenuation bands. These terminations are mid-series; that is, the entering element being one-half of the normal series element; or mid-shunt; that is, the entering element being twice the impedance of the normal shunt element. The corresponding impedances are called the mid-series and mid-shunt characteristic or iterative impedances.

If we retain the first few sections of such a structure and terminate them with a resistance which is equal to the resistance impedance of the infinite line from which they were taken, the characteristics are substantially unchanged. It is understood, of course, that this resistance equals approximately the resistance impedance of the remainder of the infinite line at most of the frequencies in the transmission band in which we are interested.

The presence of small amounts of damping in the various elements also has but slight effect on the general characteristics. These results could in general be readily applied to the various telephone transmission problems because the source and load between which the filter system was inserted generally had or could be made to have a resistance impedance nearly equalling the mid-series or mid-shunt impedance of the filter within the transmission band. The filter and terminating impedances may then be said to be matched. Where adjacent sections in the filter have impedances similar in character but different in absolute magnitude they may be joined by a suitable transformer.

Many early attempts were made to design mechanical transmission systems having a wide frequency range in which highly damped single or multi-resonant systems were employed. In these attempts both of the obvious methods of increasing the damping were used, namely, that of adding a resistance to the system and that of increasing the value of the compliance and decreasing mass in such proportion as to maintain the same natural frequency. The former of these methods reduces the sensitivity of the system at the point where it is most efficient (See Fig. 9), while the second method increases the response at the points where the system is less sensitive, namely, away from its resonance point. Fig. 9 shows four curves—first, a singly resonant system, Curve A; second, the same system with friction added, Curve B; third, the same system without the added friction but with an increase in compliance and a decrease in mass such that the natural period remains the same, Curve C; and fourth, a band pass type of circuit whose resistance impedance is the same as that of the system shown in Curve A. (See Curve D.)

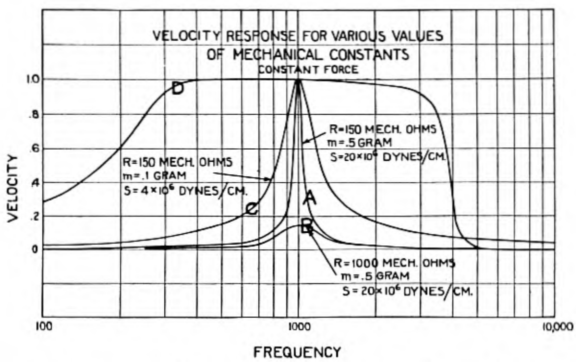


Fig. 9—Velocity response for various values of mechanical constants

The results of filter theory have shown how these resonances should be coordinated so that when a proper resistance termination is used high efficiency and equal sensitivity are obtained over a definite band of frequencies by elimination of response to all frequencies outside the band. With the electrical case of a repeated filter, each section considered by itself resonates at the same frequency but when combined into a short-circuited filter of n sections, there will be n natural frequencies. However, when such a system is terminated with a resistance which equals the nominal characteristic impedance in the transmission band, uniform response in the terminating resistance is obtained over the entire band.

Detailed Analysis of Mechanical and Electrical Analogs [7]

Before going on with a detailed treatment of the electrical analogs of the mechanical structures used in the problem of phonographic reproduction, a list of the corresponding quantities used in the two systems will be given, together with the symbols employed.

Mechanical		Electrical	
Force	$= F$ (dynes)	Voltage	$= E$ (volts)
Velocity	$= v$ (cm./sec.)	Current	$= i$ (amperes)
Displacement	$= s$ (cm.)	Charge	$= q$ (coulombs)
Impedance or mechanical ohms	$= z$ (dyne sec./cm.)	Impedance	$= Z$ (ohms)
Resistance	$= r$ (dyne sec./cm.)	Resistance	$= R$ (ohms)
Reactance	$= x$ (dyne sec./cm.)	Reactance	$= X$ (ohms)
Mass	$= m$ (gms.)	Inductance	$= L$ (henries)
Compliance	$= c$ (cm./dyne)[8]	Capacity	$= C$ (farads)

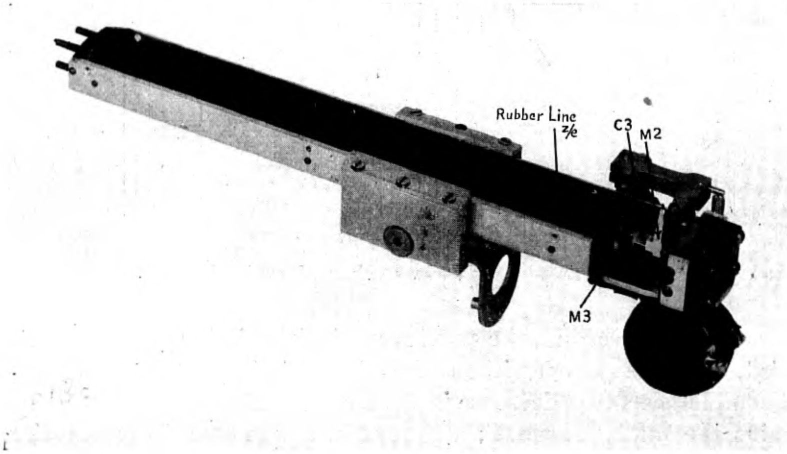


Fig. 10—This figure shows an electromagnetic recorder complete except for the bottom of the case

In addition to the above certain other quantities such as angular displacement, pressure and impedance per unit area, and a few others which have no direct electrical analog will be used. These quantities,

[7] The authors wish to express their appreciation to Mr. E. L. Norton for his courtesy in working out the mathematics of the mechanical and electrical analogs which are shown in this paper.

[8] H. W. Nichols, "Theory of Variable Dynamical Electrical Systems," *Phys. Rev.* Vol. 10, 1917.

however, are either standard in the literature or may always be reduced to those given above.

As illustrations of the general methods employed certain important portions of the reproducer will be considered in detail. Considering first the electrical analog of the air chamber [9] between the diaphragm and horn, we make use of the following list of symbols (see Figs. 3, 4, 15 and 16):

m_3 = Effective mass of diaphragm in grams;
A_1 = Equivalent area of diaphragm in cms^2;
c_6 = Compliance of edge of diaphragm;
c_7 = Compliance of air chamber;
A_2 = Area of throat of horn;
z_h = Impedance of horn—Vector ratio of applied force at the throat of the horn to the resultant linear velocity of the air;
s_1 = Displacement of diaphragm;
v_1 = Velocity of diaphragm;
s_2 = Displacement of air in throat of horn;
v_2 = Velocity of air in throat of horn;
P_0 and V_0 = Initial pressure and volume of air-chamber;
F = Force applied to diaphragm;
p = Small change of pressure in air-chamber.

For a small change p in the pressure within the air-chamber we have:

$$p=\frac{n(A_1s_1-A_2s_2)P_0}{V_0} \tag{1}$$

where $n=1$ for an isothermal change and 1.4 for an adiabatic change. For the case under consideration $n=1.4$ very nearly.

If the horn opening is closed, $s_2=0$, and we get for the compliance of the air chamber as measured from the diaphragm

$$c_7=\frac{s_1}{pA_1}=\frac{V_0}{np_0A_1^2}.$$

We have the two force equations

$$m_3\frac{dv_1}{dt}+\frac{s_1}{c_6}+pA_1=F \tag{2}$$

$$z_hv_2-pA_2=0 \tag{3}$$

[9] The use of the air chamber to increase the loading effect of the horn on the diaphragm has been appreciated for a number of years. It has been used in telephone receivers, phonographs, and loud speaking receivers since their earliest developments. A treatment of the force equations of the air-chamber was given by Hanna & Slepian, "The Function and Design of Horns for Loud Speakers." Trans. *A. I. E. E.*, 1924, p. 393. The equivalent structure, however, was analysed as a compliance and resistance in series instead of in shunt.

or substituting the values of p and c_7

$$m_3\frac{dv_1}{dt}+\frac{s_1}{c_6}+\frac{1}{c_7}\left[s_1-\left(\frac{A_2}{A_1}\right)s_2=F\right] \tag{4}$$

$$z_hv_2+\frac{1}{c_7}\left[\left(\frac{A_2}{A_1}\right)^2s_2-\left(\frac{A_2}{A_1}\right)s_1\right]=0 \tag{5}$$

If $v_1=j\omega s_1$, etc.

$$z_1v_1-z_mv_2=F$$
$$z_2v_2-z_mv_1=0$$

where

$$z_1=j\left(\omega m_3-\frac{1}{\omega c_6}-\frac{1}{\omega c_7}\right),$$

$$z_2=\left[z_h-j\left(\frac{A_2}{A_1}\right)^2\frac{1}{\omega c_7}\right],$$

$$z_m=-j\left(\frac{A_2}{A_1}\right)\frac{1}{\omega c_7}\cdot$$

Considering now the analogous electrical circuit, and assuming the velocity, current, force and voltage to vary sinusoidally, we have the parallel relationship for the steady state conditions:

$$L_3\frac{di_1}{dt}+\frac{q_1}{C_6}+\frac{1}{C_7}\left[q_1-\left(\frac{N_2}{N_1}\right)q_2\right]=E,$$

$$Z_hi_2+\frac{1}{C_7}\left[\left(\frac{N_2}{N_1}\right)^2q_2-\left(\frac{N_2}{N_1}\right)q_1=0.\right.$$

where $\dfrac{N_2}{N_1}$ = turns ratio of ideal transformer (Fig. 4).

If $i_1=\mathrm{j}\omega q_1$, etc.

$$Z_1i_1-Z_mi_2=E$$
$$Z_2i_2-Z_mi_1=0$$

where

$$Z_1=j\left(\omega L_3-\frac{1}{\omega C_6}-\frac{1}{\omega C_7}\right),$$

$$Z_2=\left[Z_h-\mathrm{j}\left(\frac{N_2}{N_1}\right)^2\frac{1}{\omega C_7}\right],$$

$$Z_m=-j\left(\frac{N_2}{N_1}\right)\frac{1}{\omega C_7}\cdot$$

The last five equations in each case give the complete solution of the network. By analogy between the two sets of equations, therefore, the air-chamber corresponding in the electrical case to a shunt capacity, c_7 is spoken of as a shunt compliance, $c_7 = \frac{V_0}{nP_0A_1^2}$, together with a transformer inserted before the horn, which has an equivalent turns ratio equal to the ratio of the areas of the diaphragm and horn openings.

Taking up now the somewhat different illustration of the needle arm, the following symbols are needed (Figs. 5, 6, 15, 16):

l_1 = Distance from pivot point to end of needle;
l_2 = Distance from pivot point to center of "spider" (Fig. 15);
I = Moment of inertia of needle arm;
m_1 = Apparent or equivalent mass of arm as measured from the center of the spider

$$= \frac{I}{l_2^2};$$

c_1 = Compliance of needle point;
c_2 = Compliance of bearing to turning of the needle arm, as measured from end of arm at the spider;
c_3 = Compliance of end of needle arm attached to spider;
s_1 = Displacement of tip of needle;
s_2 = Displacement of end of arm at the spider;
s_3 = Displacement of spider;
z_s = Mechanical impedance of spider and remainder of structure = Vector ratio of applied force to resultant velocity;
θ = Angular displacement of needle arm;
F = Applied force at needle point.

We have the three force equations:

$$\frac{s_1 - l_1\theta}{c_1} = F \tag{6}$$

$$I\frac{d^2\theta}{dt^2} + \frac{(l_1\theta - s_1)l_1}{c_1} + \frac{\theta l_2^2}{c_2} + \frac{(l_2\theta - s_3)l_2}{c_3} = 0 \tag{7}$$

$$\frac{s_3 - l_2\theta}{c_3} + z_s\frac{ds_3}{dt} = 0 \tag{8}$$

Replacing θ by $\frac{s_2}{l_2}$ and I by $m_1 l_2^2$ gives:

$$\frac{s_1 - \frac{l_1}{l_2} s_2}{c_1} = F \tag{9}$$

$$m_1 \frac{d^2 s_2}{dt^2} + s_2 \left[\left(\frac{l_1}{l_2} \right)^2 \frac{1}{c_1} + \frac{1}{c_2} + \frac{1}{c_3} \right] - \frac{l_1}{l_2} \frac{s_1}{c_1} - \frac{s_3}{c_3} = 0 \tag{10}$$

$$\frac{s_3 - s_2}{c_3} + z_s \frac{ds_3}{dt} = 0 \tag{11}$$

Considering now the parallel mechanical electrical circuits, and assuming as before sine functions for v, i, F, and E, we have:

Mechanical Case, substituting $v_1 = j\, \omega\, s_1$, etc., in the last equations:

$$-j \frac{v_1}{\omega c_1} + j \frac{l_1}{l_2} \frac{v_2}{\omega c_1} = F,$$

$$j v_2 \left[\omega m_1 - \left(\frac{l_1}{l_2} \right)^2 \frac{1}{\omega c_1} - \frac{1}{\omega c_2} - \frac{1}{\omega c_3} \right]$$

$$+ j \frac{l_1}{l_2} \frac{v_1}{\omega c_1} + j \frac{v_3}{\omega c_3} = 0,$$

$$j \frac{v_2}{\omega c_3} + v_3 \left(z_s - j \frac{1}{\omega c_3} \right) = 0.$$

Electrical Case, with ideal transformer of turns ratio $\frac{N_2}{N_1}$:

$$-j \frac{i_1}{\omega C_1} + j \left(\frac{N_2}{N_1} \right) \frac{i_2}{\omega C_1} = E,$$

$$j i_2 \left[\omega L_1 - \left(\frac{N_2}{N_1} \right)^2 \frac{1}{\omega C_1} - \frac{1}{\omega C_2} - \frac{1}{\omega C_3} \right] + j \left(\frac{N_2}{N_1} \right) \frac{i_1}{\omega C_1} + j \frac{i_3}{\omega C_3} = 0,$$

$$j \frac{i_2}{\omega C_3} + i_3 \left(Z_s - j \frac{1}{\omega C_3} \right) = 0.$$

The analogy between the two sets of equations is quite obvious. It will be noticed that the effect of the lever arm is to introduce an equivalent transformer of a turns ratio which is the reciprocal of the corresponding lengths of the arms.

The general method of deducing the equivalent electric circuits should be clear from the above illustrations of the air-chamber and of the needle arm. For example, in the spider section, Fig. 15, the mass is driven directly by the force from the needle-arm compliance, there being a small series compliance in the connection owing to bending of connecting rod. The diaphragm is connected through the

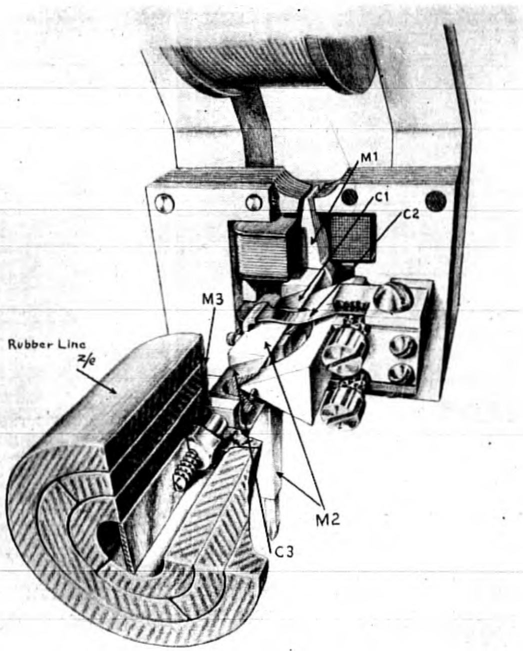


Fig. 11—Detailed drawing of the mechanical filter of an electromagnetic recorder. (Lettering same as in Fig. 12)

compliance of the prongs of the spider. The equivalent circuits are shown in Figs. 7 and 16.

The equations of this network may be obtained from the equations for the needle arm by placing c_1 equal to zero, taking a unity ratio transformer, and substituting m_2 for m_1, c_4 for c_2, c_5 for c_3 and z_d for z_s.

Another type of network which occurs frequently in the building of mechanical vibrating systems is represented diagrammatically in Fig. 8. This is clearly a particular case of Fig. 7 with c_4 made infinite.

By combining Fig. 6 representing the needle arm, Fig. 7, representing the spider section and Fig. 4 representing the diaphragm, air-chamber and horn, the complete reproducer may be built up. The resultant network is shown in Fig. 16. Since methods are available in the theory of electric wave filters to determine the proper values of the elements of the complete network for a free transfer of energy throughout an assigned frequency band, the analogous mechanical elements may be determined in the same manner.

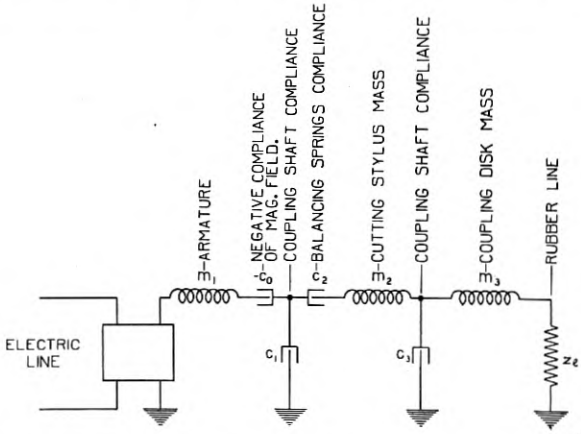


Fig. 12—Equivalent electric circuit of the electromagnetic recorder

General Design of Mechanical Systems

In designing mechanical systems of the band pass type, the problem is three fold—first, that of arranging the masses and compliances such that they form repeated filter sections; second, determining the magnitude of these quantities so that with or without transformers the separate sections all have the same cut-off frequenceis[10] and characteristic impedances; third, to provide the proper resistance termination. Where the transmitted mechanical power has not been radiated as sound this third part has been one of the most difficult to fulfill.

In designing these systems, practical difficulties arose—first, the difficulty of insuring that the parts vibrated in the desired degrees of freedom only, and second, the difficulty of determining the magnitudes of the various effective masses, compliances and resistances. Before the work to be described could be carried out practically it became necessary to develop a method of measuring mechanical impedances[11].

[10] It is of course permissible to have a section having a higher cut-off than the others provided its characteristic impedance is the same as that of the others over the transmission band of those having the lower cut-off.

[11] Kennelly, A. E. and Affel, H. A., "The Mechanics of Telephone Receiver Diaphragms, as Derived from their Motional Impedance Circles," *Proc. A. A. A. S.*, Vol. 51, No. 8, November, 1915.
Kennelly, A. E. and Pierce, G. W., "The Impedance of Telephone Receivers as Affected by the Motion of their Diaphragms," *Proc. A. A. A. S.*, Vol. 48, No. 6, September, 1912.

Such a method has been developed which at the present time covers a range of frequencies from somewhere below 50 to about 4,500 pps. Work is still being continued to extend the method to the higher frequencies. This method of measurement has been very useful not

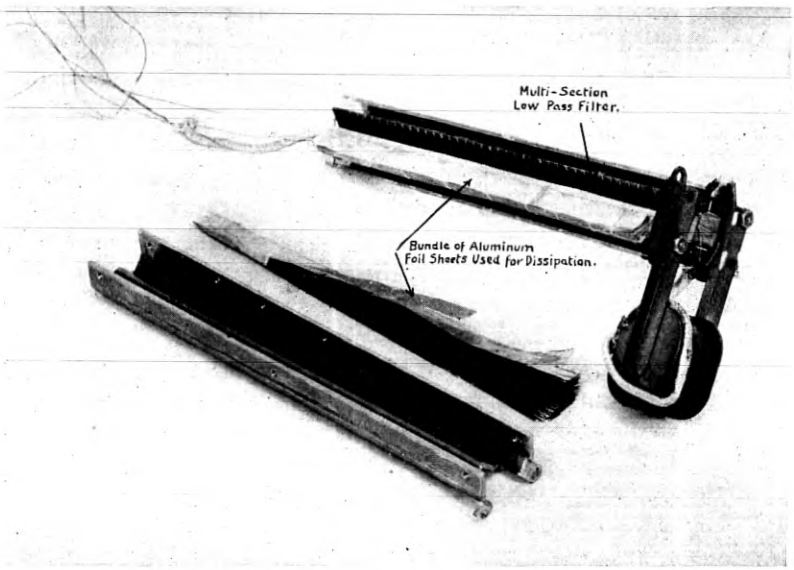


Fig. 13—Electromagnetic recorder using lumped loaded termination
The method of furnishing dissipation to the lumped loaded line is shown

only in determining the magnitudes of the impedances in the degrees of freedom in which it is desired that they shall operate, but in determining the impedances to motion of the various parts in directions in which they should not be permitted to vibrate. In connection with the measurement of the magnitudes of the parts in the desired degrees of freedom this method enables us to determine the constants of the mechanical networks under their conditions of operation. Experience so far has indicated that when all the degrees of freedom have been taken into account and when the dynamic axes of vibration have been properly chosen, the static and dynamic constants of the parts are the same, and it is then possible to check the parts by simple static measurements. In the early attempts to build these systems very large discrepancies between the static and dynamic characteristics were found.

The Recorder

One of the early practical phonographic applications of electric filter design to mechanical problems was the development of an electromagnetic recorder. The instrument as finally constructed is essentially a properly terminated three-section mechanical filter in which the recording stylus and its holder constitute the series mass in the second section. Since a filter of this type appears at its input end as approximately a pure resistance within the transmission band, the current in the series inductances, that is, in the mechanical case, the velocity of the series masses is proportional to the driving force.

Figs. 10, 11 and 12 show respectively, a complete recorder, a drawing of the mechanical filter of such a recorder and a diagram of the equivalent electric circuit. The armature acts as the series mass m_1 in the first section; the magnetic field as the series negative compliance, $-c_0$; the shaft between the armature and the stylus holder as the shunt compliance c_1; the balancing springs as the series compliance c_2; the stylus holder and the stylus as the series mass m_2; the shaft between the stylus holder and the disk, coupling the system to the terminating resistance, as the compliance c_3; the coupling disk as the series mass m_3 and the terminating line as approximately a mechanical resistance.

All of these equivalents are seen from the simple analog previously outlined with the exception of the terminating resistance and the negative compliance, $-c_0$. The terminating resistance was originally made up of a series of filter sections of lumped series masses and shunt compliances with a small amount of damping added to the motion of each of the series masses. Fig. 13 shows one of the early recorders equipped with this type of resistance termination. The reason for using such a complicated termination lies in the fact that most of the known mechanical resistances have values which are functions of frequency or of amplitude or both. Also in most cases, the mechanical resistance is accompanied by either a mass or compliance reactance. By using a multi-section filter which is sufficiently long so that a wave entering it will be essentially absorbed before it has reached the far end, been reflected and returned to the entering end, it has been possible to use imperfect types of damping for this line and still obtain over the desired band, an essentially pure resistance at the input end.

More recently a continuous line has been developed which is much easier of practical attainment than the complicated lump-loaded filter. The recorder shown in Fig. 10 is so equipped.

Fig. 14 shows calibration curves of three types of recorders. The bottom curve shows an early type of highly damped singly resonant system. The middle curve is a calibration of a low pass mechanical filter type using lumped loading in the resistance line. The upper curve shows the calibration of the recorder shown in Fig. 10.

The compliance $-c_0$ is a mechanical quantity for which there is no simple electric analog. In a balanced armature type of structure such as that shown in Fig. 11, the action of the field on the armature, when it is at its center point, is balanced. If, however, the armature be de-

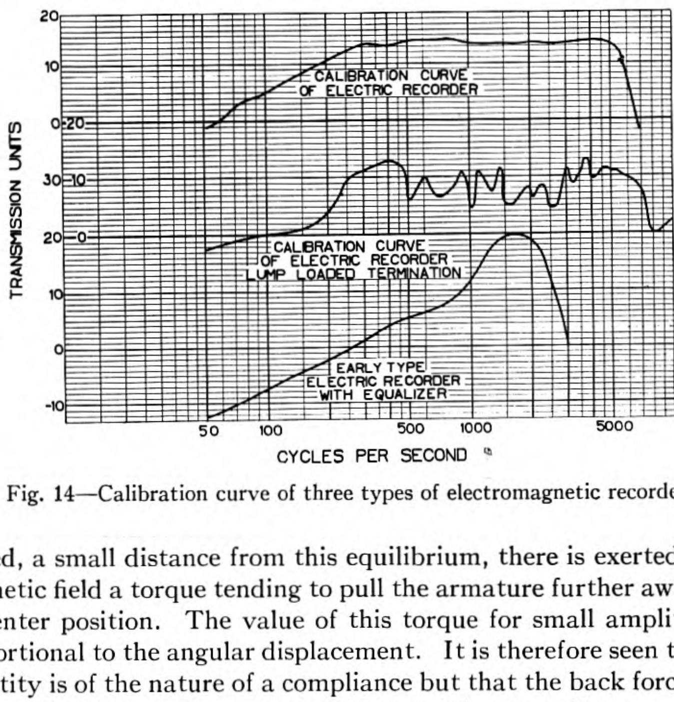


Fig. 14—Calibration curve of three types of electromagnetic recorders

flected, a small distance from this equilibrium, there is exerted by the magnetic field a torque tending to pull the armature further away from its center position. The value of this torque for small amplitudes is proportional to the angular displacement. It is therefore seen that this quantity is of the nature of a compliance but that the back force is in a reverse direction to that required for a positive compliance.

Design of the Reproducing Apparatus

As the analogy between the mechanical and electrical filter is more perfectly shown in the case of the reproducing equipment, its detailed quantitative description will now be given. Figs. 15 and 16 show respectively a diagram of the reproducing system and its equivalent electric circuit. From these diagrams it is evident which units in the mechanical system correspond to the various electrical parts. As the series compliances c_2, c_4 and c_6 have been made so large that the low frequency cut-off caused by them lies well below the low frequency cut-off of the horn, an inappreciable error is introduced in

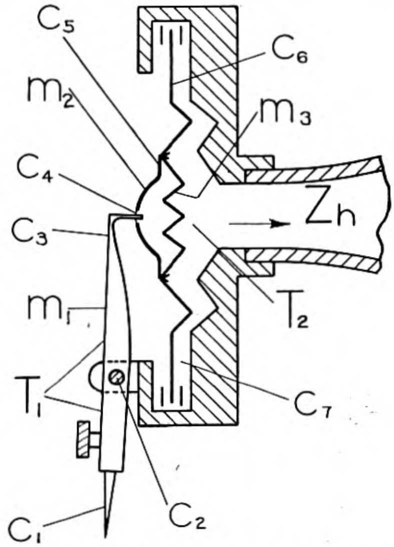


Fig. 15—Diagrammatic sketch of the mechanical system of the phonograph

using for design purposes formulas of low pass filters[12]. The two formulas which will be used are as follows:

$$f_c = \frac{1}{\pi}\sqrt{\frac{1}{mc}} \tag{12}$$

Where

f_c = cut-off frequency of a lumped transmission system in cycles per second

c = shunt compliance per section in centimeters per dynes

m = series mass per section in grams

$$z_o = \sqrt{\frac{m}{c}} \tag{13}$$

where z_0[13] is the value of characteristic impedance over the greater part of the band range.

[12] Campbell, G. A., "On loaded lines in Telephonic Transmission," *Phil. Mag.*, March, 1903.

[13] z_0 may be called nominal mid-shunt or mid-series impedance. Their actual values in the transmission band being at any frequency f.,

$$\text{mid-series} = z_0\sqrt{1-\left(\frac{f}{f_c}\right)^2} \qquad \text{mid-shunt} = \frac{z^0}{\sqrt{1-\left(\frac{f}{f_c}\right)^2}}$$

Equations (12) and (13) which form the basis of the design work contain four variables, f_c, c, m and z_0. It is, therefore, necessary to determine two of them by the physical requirements of the problem after which the other two are determined. The upper cut-off frequency f_c was arbitrarily chosen at 5000 pps. as a compromise between the highest frequency occurring on the record and the increase in surface noise

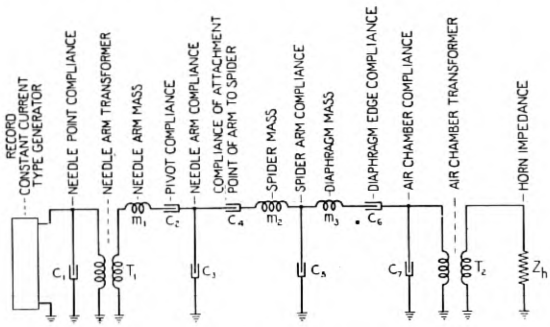


Fig. 16—Electric equivalent of the system shown in Fig. 15

as the cut-off is raised. The choice of the other arbitrarily set variable came after considerable preliminary experimenting and was fixed by the difficulty of obtaining a diaphragm which is light enough and has a large enough area. Hence the effective mass of the diaphragm m_3, (Figs. 15–16) was fixed at 0.186 grams which value can be obtained by careful design. The effective area can be made as large as 13 square centimeters. For convenience let the arbitrary value chosen for $f_c = \bar{f_c}$ and the value of $m = \overline{m}_3$.

Solving Equations (12) and (13) for c and z_0, we get

$$c = \frac{1}{\pi^2 \bar{f_c}^2 \overline{m}_3}, \tag{14}$$

$$z_0 = \pi \bar{f_c} \overline{m}_3; \tag{15}$$

also

$$z_0 = \frac{1}{\pi c \bar{f_c}} \tag{16}$$

In order to obtain the low value of mass mentioned, with a large enough area, it was necessary to make the diaphragm of a very stiff light material. An aluminum alloy sheet 0.0017 in. thick was chosen and concentrically corrugated as shown in Figs. 17 and 18. These corrugations are spaced sufficiently close so that the natural periods of the flat surfaces are all above $\bar{f}_c$. To insure that this central stiffened portion should vibrate with approximate plunger action, which is more efficient than diaphragm action, it is driven at six points near its periphery.

Reference to Figs. 15 and 16 and Equation (14) shows that the compliance of the air chamber c_7, of the spider legs c_5 and shunt tip of the needle arm c_3 are determined. Also the mass of the spider m_2 and the effective mass of the needle arm m_1, as viewed at the point where it is attached to the spider, are determined.

The impedance looking into the system from the record is determined by the rate at which it is necessary to radiate energy in order that the reproduction may be loud enough. The power taken from the record is approximately $v^2 z_0$ since z_0 is a resistance over most of the band. Experiment has shown this value of z_0 to be approximately 4500 mechanical ohms.

But substituting in Equation (13) the value of $\overline{m}_3$, and from Equation (14) the value of c_5, we find that the impedance is only 2920 mechanical ohms. It is, therefore, necessary to use a transformer whose impedance ratio is $\frac{4500}{2920}$. From this and a knowledge of filter structures the needle-point compliance can be determined. The value obtained is easily realized with commercial types of needle.

It will be noted that the record is shown in Fig. 16 as a constant current generator, *i. e.*, a generator whose impedance appears high as viewed from the needle point. That this is necessary is obvious when it is remembered that, if the impedance looking back into the record were to equal the impedance of the filter system, the walls of the record would have to yield an amount comparable with one-half the amplitude of the lateral cut. This would cause a breakdown of the record material with consequent damage.

The design of the system is, therefore, complete except for the resistance termination which is supplied by the horn for all frequencies above its low frequency cut-off. The characteristics of the horn will be dealt with later. The resistance within the band looking in at the small end of the horn is $G A_2$ where G equals the mechanical ohms per square centimeter of an infinite cylindrical tube of the same area, and A_2 equals the area in square centimeters of the small end of the horn.

Let A_1 = the effective plunger area of the diaphragm (as previously mentioned this is 13 sq. cm.). The impedance looking back at the diaphragm is

$$z_0 = \pi \bar{f}_c \overline{m}_3 = 2920 \text{ mechanical ohms}$$

from Equation (15), and the impedance looking at a horn whose small end area equals A_2 is

$$z_h = r_0 = A_2 G \tag{17}$$

Substituting

$$A_2 = 13 \text{ sq. cm.}$$
$$G = 41 \text{ ohms per cm.}^2$$

we get

$$z_h = r_0 = 533 \text{ mechanical ohms}$$

This is entirely insufficient so that the air-chamber transformer becomes necessary.

To calculate the necessary ratio of areas on the two sides of the air-chamber transformer, the following formula is needed. The formula assumes the chamber to be relatively small compared with all wave lengths of the sound to be transmitted, that is, the pressure changes throughout the chamber are substantially in phase.

$$\frac{z_0}{z_h} = \left(\frac{v_2}{v_1}\right)^2 = \left(\frac{F_1}{F_2}\right)^2 = \left(\frac{A_1}{A_2}\right)^2 \tag{18}$$

where

z_0 = the impedance of the primary side of the transformer in mechanical ohms;

z_h = the impedance on the secondary side of the transformer in mechanical ohms, *i.e.*, the horn impedance;

v_1 = mechanical current, *i.e.*, velocity on the primary side of the transformer in centimeters per second;

v_2 = mechanical current on the secondary side of the transformer in centimeters per second;

F_1 = alternating force on primary side of air-chamber transformer in dynes;

F_2 = alternating force on secondary side of air-chamber transformer in dynes;

A_1 = effective area working into the primary side of the air-chamber in centimeters squared;

A_2 = effective area working into the secondary side of the air-chamber in centimeters squared.

The characteristic impedance of the line on the diaphragm or primary side of the air-chamber as shown by equation (15) is

$$z_0 = \pi \bar{f}_c \bar{m}_3. \tag{19}$$

From Equation (17) the characteristic impedance on the horn or secondary side is

$$z_h = GA_2. \tag{20}$$

Therefore,

$$\left(\frac{A_2}{A_1}\right)^2=\frac{z_h}{z_0}=\frac{GA_2}{\pi f_c m_3} \tag{21}$$

and solving this for A_2, we get

$$A_2=\frac{GA_1^2}{\pi f_c m_3} \tag{22}$$

The equivalence of the air-chamber to a transformer shunted by a compliance is shown earlier in the paper.

In applying the foregoing method of design to a practical structure, a number of design problems had to be solved. The construction of the diaphragm and the method by which it is actuated have been already described, except for the tangential corrugations constituting the series compliance. The use of these corrugations results in the

Fig. 17—Photograph of mechanical reproducing system without the horn

value of the series compliance being practically independent of the nature of the clamping, and has eliminated a tendency to "rattle" introduced by unevenness in the clamping surfaces.

Another feature in connection with the sound box is the needle-arm bearing shown in Figs. 17 and 18. Ordinary knife-edge bearings are not sufficiently rigid as fulcrums and the rotational reactance as well as the rotational resistance is undesirably large. A construction which has been found to meet the necessary requirements is the ball bearing type with the steel balls held in position by magnetic pull. By making the ball-containing case of soft steel and magnetizing the shaft, it has been possible to manufacture this bearing reliably and cheaply.

The horn which has been used as a terminating resistance to the mechanical filter structure is a logarithmic one. The general properties of logarithmic horns have been understood for some time.[14]

There are two fundamental constants of such a horn—the first is the area of the large end and the second the rate of taper. The area of the

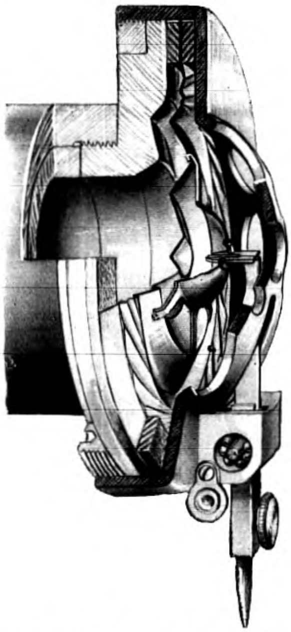


Fig. 18—Sectional drawing showing construction of the system shown in Fig. 17

mouth determines the lowest frequency which is radiated satisfactorily. The energy of the frequencies below this is largely reflected if it is permitted to reach the mouth.

From the equations given by Webster, [14] it can be shown that all logarithmic horns have a low frequency cut-off which is determined by the rate of taper. If the rate of taper is so proportioned that its resulting cut-off prevents the lower frequencies from reaching the horn mouth, the horn will then radiate all frequencies reaching its mouth and very little reflection will result.[15] It is, therefore, possible to build a horn having no marked fundamental resonance.

[14] Webster, A. G., "Acoustical Impedance and Theory of Horns and Phonograph," *Proc.* Nat. Acad. of Sci., 1919.

[15] The authors wish to express their appreciation in this connection of the work of Mr. P. B. Flanders who carried out the mathematical investigation of these relationships and to Mr. A. L. Thuras who checked experimentally the mathematical theory.

Since the characteristics of the horn are determined by the area of its mouth and by its rate of taper the length of the horn is determined by the area of the small end. This area is determined in turn by the mechanical impedance and effective area of the system which it is terminating, as shown in Equation (22). It is seen, therefore, that the length of the horn should not be considered as a fundamental constant. A paper describing the design of horns based on these principles is being prepared.

An interesting feature of the horn which has been built commercially is its method of folding. The sketch in Fig. 19 shows a shadow picture

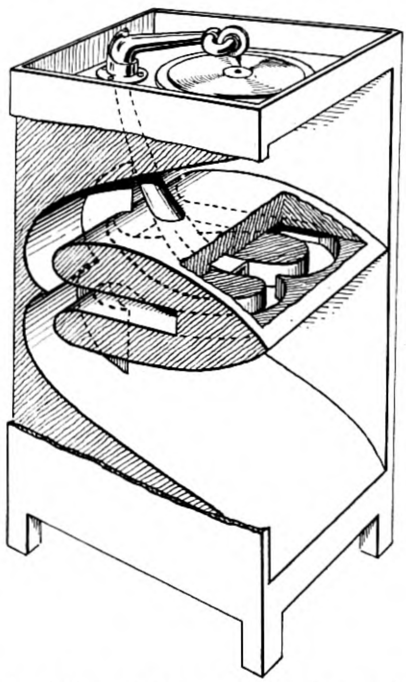


Fig. 19—Sectional view of the folded horn showing the air passage

of the horn. It will be noticed that the sound passage is folded only in its thin direction, which permits the radius of the turns to be small and thereby makes the folding compact.

Fig. 20 shows the frequency characteristic of a phonograph designed as shown above with a logarithmic horn whose rate of taper and area of mouth opening place the low cut-off at about 115 cycles. It also shows the characteristics of one of the best of the old style phonographs. Curve A represents the new machine, while Curve B represents the old style standard machine. Since the vertical scale used in this graph is logarithmic the full difference between the two instru-

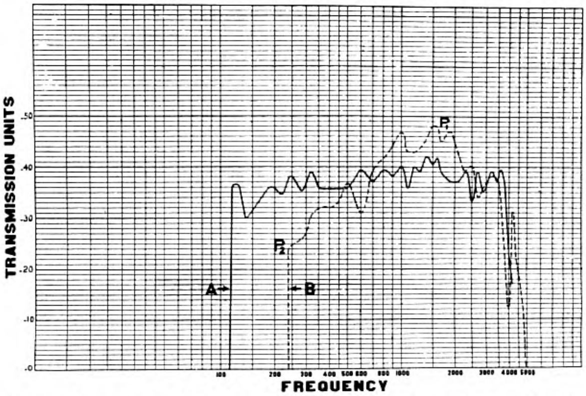


Fig. 20—Response frequency characteristic of two phonographs. Curve A shows the characteristic of the band pass filter type described. Curve B shows the characteristic of one of the best commercial machines previously on the market

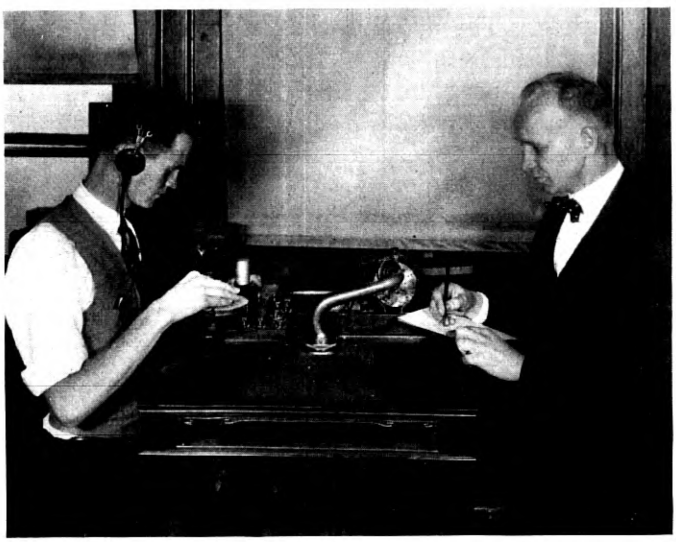


Fig. 21—Bridge for measuring mechanical impedance, being used for determining the impedance of a phonograph horn

ments does not appear at first glance. Some idea as to the magnitude of this difference can be obtained, however, by noting that the points P_1 on the curve of the old machine stands at a power level about 250 times as great as P_2.

Abstracts of Recent Technical Papers from Bell System Sources

Complex Magnetization.[1] EUGENE PETERSON. Magnetization of silicon steel by two sinusoidal fields of differing frequencies. The energy loss W per cycle and the flux density B associated with each of the two frequencies were determined when the two sinusoidal magnetizing forces were simultaneously impressed on a toroidal silicon steel core built up of one-mil laminations. A null method was used which permitted suppression of the modulated currents and constancy of the impressed currents during manipulation for balance. The frequencies used were 400, 821 and 1582. Six sets of measurements were taken with fixed magnetizing forces ranging from 0.5 to 10 gilberts/cm and superposed forces up to 15 gilberts/cm. The results show that the effect of superposition depends upon the relative amplitudes and upon the frequency ratio R of the superposed frequency to the other. At low fixed fields W and B go through maxima as the superposed field is increased, the maximum value increasing with R. The maximum is less pronounced or absent for the higher fixed fields. In general B is smaller with a low than with a high value of R other things being equal. The effect on W is not as sharply defined; in general the effect of superposition is more pronounced the higher the superposed frequency. The amplitude effect and frequency ratio effect are shown to be in general agreement with conclusions drawn from mathematical treatment of somewhat simplified cases and it is concluded that the effects are not inconsistent with purely hysteretic phenomena.

Some Photographic Problems Encountered in the Transmission of Pictures by Electricity.[1] HERBERT E. IVES. This paper considers some of the problems of photographic tone reproduction, which arise upon the introduction of an electrical transmission system between a picture placed on sending apparatus in one place and the copy of the picture made by receiving apparatus in another place. Some of these problems arise because of limitations introduced by the use of the electrical transmission line; others arise because of opportunities for the control of picture quality which are not afforded by ordinary photographic methods. As an illustration of one of these limitations

[1] *Physical Review*, Vol. 27, pp. 318–328, March, 1926.

[2] *Journal of the Optical Society of America and Review of Scientific Instruments*, March, 1926, pp. 173–194.

may be mentioned the fact that the original picture, for instance a photographic negative, is not seen by the operator at the receiving end. He cannot, therefore, by using his photographic knowledge and experience, choose printing media and decide upon conditions of exposure and development. As an illustration of the opportunities introduced by an electrical picture transmission apparatus may be noted the possibility of so poling the electrical elements that the received picture may be either a positive or negative, irrespective of the nature of the original at the sending end.

While in other picture transmission systems other problems arise peculiar to these systems, it is believed that although the questions considered are those presented in commercial operation in the Bell System, they are, to a certain extent, common to all electrical picture transmission apparatus.

A Radio Field-Strength Measuring System for Frequencies up to Forty Megacycles.[3] H. T. FRIIS and E. BRUCE. In previous types of radio field strength measurement apparatus it is very difficult to reproduce accurately the small comparison voltages at very high frequencies, due to reactive effects in the attenuating networks. The "tube voltmeter" is practically the only reliable instrument available at high frequency measurement work. New measurement sets for very high frequency signals have, therefore, been developed. The apparatus is a double detection receiving set which is equipped with a calibrated intermediate frequency attenuator and a local signal comparison oscillator. The local signal is measured by means of the intermediate frequency detector which is calibrated as a tube voltmeter and all required attenuations are made at the relatively low and fixed intermediate frequency.

A New Mechanical Test for Rubber Insulation.[4] C. L. HIPPENSTEEL. This paper discusses the development of a rapid routine test which will numerically express the ability of the rubber insulation to resist cutting by the conductor at the points of support and to resist cracking at points of extreme flexure. Up to the present time no one test of that nature has been described.

[3] Presented at a meeting of the Institute of Radio Engineers, May 5, 1926.

[4] *Industrial and Engineering Chemistry*, April, 1926.

Contributors to this Issue

C. F. Sacia, B.E.E., University of Michigan, 1916; Engineering Department of the Western Electric Company, 1916-24; Bell Telephone Laboratories, Inc., 1925—. Mr. Sacia has been engaged upon methods for recording and analyzing speech.

C. J. Beck, Western Electric Company Engineering Department, 1917; U. S. Army Air Service, 1918-19.—Since returning to the Laboratories, Mr. Beck has been associated with Mr. Sacia in speech analyses.

J. J. Gilbert, A.B., University of Pennsylvania, 1909; Harvard, 1910-11; Chicago, 1911-12; E.E., Armour Institute, 1915; instructor of electrical engineering, Armour, 1912-17; Captain Signal Corps, 1917-19; Engineering Department, Western Electric Company, 1919; Bell Telephone Laboratories, Inc., 1925—. Mr. Gilbert has worked primarily on submarine cable problems.

R. B. Shanck, B.E.E., Ohio State University, 1915. Railroad Telegraph Service, 1909-10; Plant Department, American Telephone and Telegraph Company, 1910-11 and summers 1912-14; Engineering Department, 1915-19; Department of Development and Research, 1919—. Since 1915 Mr. Shanck has been engaged in transmission development work, on telegraph systems and compositing arrangements.

F. B. Llewellyn, M.E., Stevens Institute of Technology, 1922; graduate student, Columbia University, 1924-5; Research Department, Western Electric Company, 1922-5; Bell Telephone Laboratories, 1925—.

Karl K. Darrow, S.B., University of Chicago, 1911; University of Paris, 1911-12; University of Berlin, 1912; Ph.D., in physics and mathematics, University of Chicago, 1917; Engineering Department Western Electric Company, 1917-24; Bell Telephone Laboratories, Inc., 1925—. Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

Joseph P. Maxfield, S.B., Massachusetts Institute of Technology, 1910; instructor in electro-chemistry, Massachusetts Institute of

Technology, 1910-12; instructor in physics, Massachusetts Institute of Technology, 1912-14; industrial research work, 1910-14; Western Electric Company, 1914-24; Bell Telephone Laboratories, 1925—. Mr. Maxfield has engaged in fundamental studies of microphonic contacts and electro-acoustic devices. Recently these activities have been directed toward the development of public address systems and methods of recording and reproducing music and speech particularly as regards their application to the phonograph and audible motion pictures.

Henry C. Harrison, A.B., Colorado College, 1910; S.B., Massachusetts Institute of Technology, 1913; instructor in electrical engineering, Massachusetts Institute of Technology, 1913-14; Western Electric Company, 1914-24; Bell Telephone Laboratories, 1925—. Mr. Harrison has made fundamental studies of receivers and carbon button microphones. More recently, his work has been principally concerned with the design and development of various mechanical apparatus to embody the principles of electric transmission theory.