THE JOURNAL
OF THE
ACOUSTICAL SOCIETY OF AMERICA

Volume 22 Number 2

MARCH • 1950

The Perception of Speech and Its Relation to Telephony

HARVEY FLETCHER AND ROGERS H. GALT
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received December 3, 1949)

The intelligibility of the speech received over a communication system is usually expressed in terms of one or another measure such as the vowel or the consonant articulation, the average speech sound articulation, the syllable articulation, the word articulation, or the sentence intelligibility. The present paper establishes relationships among several of these measures and the articulation index. Relationships based upon statistical considerations are compared with the results of observations. Functions are developed which permit the calculation of articulation index and hence of articulation for communication systems which include a wide variety of response *versus* frequency characteristics and of noise conditions, as well as several special types of distortion. Although the treatment is predominantly empirical, the functions and processes are closely related to various fundamental properties of speech and hearing. Four principal series of articulation tests are cited in detail, some of which have been described in published articles by various persons. The response and the noise, if any, are given for each of these cases and the observed articulations are compared with values calculated by the method here presented. The application of the computational method to the perception of speech by deafened persons is reserved for a subsequent paper. A "Foreword" to the present paper describes the historical importance of articulation tests in the Bell Telephone System.

CONTENTS

FOREWORD

THIRTY years ago the instruments used in commercial telephone systems had peaked response characteristics, and distortion and noise were not well controlled. Laboratory tests were devised whereby the loss of intelligibility of the received speech could be measured for each degrading factor separately. These tests, generally known as articulation tests, were extended and improved over a considerable period of years, and they played an important part in the development of the modern telephone instruments and circuits which now have quite uniform response characteristics and are relatively free from distortion and noise. During the recent war, articulation tests again became prominent in studies leading to improved systems of communication in situations typical of warfare and particularly under circumstances of intense noise. The results of a wide variety of such articulation tests, made in the years 1919 to 1945, are brought together and coordinated in the present paper, and a method is described for calculating the results which over the years have been obtained through elaborate and expensive testing.

Today in the Bell System in testing instrumentalities intended for the commercial telephone plant, articulation testing has become a secondary procedure in the evaluation of transmission performance and is employed only occasionally and in an abbreviated fashion as compared with the practice in earlier years. More emphasis is now placed upon loudness, naturalness, annoyance effects, and the effort required by the subscriber to use the telephone. Transmission rating procedures are based upon field measurements such as repetition observations and volume indicator measurements on actual working circuits as discussed in papers as that of W. H. Martin[1] and that of F. W. McKown and J. W. Emling.[2] Nevertheless, questions arise and doubtless will continue to arise involving a knowledge of the harm or benefit to the intelligibility of the received speech when some change is made in instruments or circuits, and in answering these questions a method of calculating articulation is of value to the telephone engineer.

Articulation Test Record

Date 3-16-28
Title of test Practice tests
Test no. 10
List Nos. 5-9-37

Syllable articulation 0.515=S
Condition tested 1500~ low pass filter
Observer W.H.S.
Caller E.B.

No.		Observed	Called	Observed	Called	Observed	Called
1	The first group is	ma′v	na′v	po′z	po′t′h	kōb √	kōb
2	Can you hear	pōch √	pōch	nēz	nēzh	shēt′h	siz
3	I will now say	seng √	seng	jo′ch √	jo′ch	fūch √	fūch
4	As the fourth write	chūd √	chūd	t′ha′m √	t′ha′m	thōl √	thōl
5	Write down	run √	run	hab √	hab	pot′h √	pot′h
6	Did you understand	chiz	kiz	def	doth	wa′m √	wa′m
7	I continue with	foz	fozh	chech	chej	gūm	gūn
8	These sounds are	lo′l √	lo′l	lun	lon	nāsh	nāth
9	Try the combination	jās	zhāth	shāl √	shāl	vo′g √	vo′g
10	Please record	t′ha′th	t′ha′sh	muz √	muz	lung	long
11	Write the following	wūr √	wūr	lēd	bēd	diz	dizh
12	Now try	yāp √	yāp	wif √	wif	kak	tak
13	Thirteen will be	mad	maj	gōst √	gōst	t′ha′r	zha′r
14	You should observe	bēch	bēk	thāv	sāv	must √	must
15	Write clearly	gēm	dēm	kōf √	kōf	yo′d √	yo′d
16	Number 16 is	t′heb	veb	ra′g √	ra′g	jet √	jet
17	You may perceive	jok	jost	thip √	thip	rēp	rēj
18	I am about to say	gaf √	gaf	yar √	yar	t′hēp	hēp
19	Try to hear	hus √	hus	zhūt √	zhūt	—	chuv
20	Please write	hiv	thit′h	kūk	tūk	t′hef	t′hesh
21	Listen carefully to	tōg √	tōg	fung √	fung	bās √	bās
22	The last group is	sho′t √	sho′t	t′hev	vesh	t′ho′f	shaf

$v=0.909$
$c=0.735$
$s=0.793$

$cvc=0.491$
$s^3=0.499$

FIG. 1. Sample articulation test record.

In the title of the present paper, the word "telephony" is used in a broad sense and is not restricted to the commercial telephone. In this paper, a telephone system signifies any talker-listener combination. Thus the method of calculation is applicable not only to commercial telephones but also to local intercommunication systems, public address systems, hearing aids, speech recording and reproducing systems, and air paths without or with reverberation and noise.

Various methods of calculating articulation have been developed in the past, each being more or less successful within a limited range of conditions. Of the methods heretofore developed at the Bell Telephone Laboratories, probably the most successful was that described by French and Steinberg.[3] The method of calculating articulation described in the present paper is based upon a much wider range of telephone systems and noise conditions than any previous method developed by the Laboratories. This method calculates correctly, within the observational error, for all of the noiseless systems included in the three large groups of articulation tests made at the Laboratories in the period from 1919–1935, and for the systems with moderate noise levels tested in 1936–1937. The method also calculates correctly for the intense pure tone interference tests made at the Laboratories, and for the tests made at the Psycho-Acoustic Laboratory at Harvard University during the recent war, involving intense noises similar to those encountered in military situations, as reported by Egan and Wiener.[4] In addition, the paper correlates a variety of types of articulation tests and gives certain previously known functions in revised forms having a more extensive basis in experiment. Thus it is hoped that the paper will be useful not only to communication engineers but also in schools where the principles of speech transmission and interpretation are being presented.

1. INTRODUCTION

There are several aspects to the perception of speech. First and foremost is the process which enables one to interpret correctly and to repeat or to record the sounds which are spoken—the interpretation aspect. Second,

[1] W. H. Martin, "Rating the transmission performance of telephone circuits," Bell Sys. Tech. J. 10, 116 (1931).

[2] F. W. McKown and J. W. Emling, "A system of effective transmission data for rating telephone circuits," Bell Sys. Tech. J. 12, 331 (1933).

[3] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," J. Acous. Soc. Am. 19, 90 (1947).

[4] J. P. Egan and F. M. Wiener, "On the articulation efficiency of bands of speech in noise," Department of Commerce, Office of Technical Service Report PB-97057 (May 1, 1945); J. P. Egan and F. M. Wiener, "On the intelligibility of bands of speech in noise," J. Acous. Soc. Am. 18, 435 (1946).

TABLE I. Groups of articulation tests with corresponding symbols for articulation.

Designation of telephone systems	Dates of articulation tests	Tests made by	Received speech recorded by	Composition of syllables: Consonants and vowels	Percent	Designation of average articulation: Speech sounds	Syllables
I	(1919–1920) (1924–1925)	BTL	Writing	(CV (VC (CVC	10) 10) 80)	s_{23}	S_{23}
II	1928–1929	BTL	Writing	CVC	100	s_3	S_3
III	(1935–1936) (1936–1937)	BTL	Machine	CVC	100	s_{3M}	S_{3M}
H	1944–1945	Harvard Univ.	Writing	(CV (VC	50) 50)	s_2	S_2

one can determine whether the heard sound is loud or soft—the loudness aspect. Third, one can determine whether the pitch is high or low—the pitch aspect. Fourth, one can determine the quality of the voice of the speaker, whether it is a child's voice, a woman's voice, or a man's voice, or whether the voice is harsh or pleasing—the quality aspect.

This paper deals with the interpretation aspect and how it is affected when speech is transmitted through various kinds of telephone systems. The method of measuring this aspect of perception is to have a speaker read aloud a certain number of speech sounds to a listener who writes what he thinks he hears. In order that the sounds may be used as they occur in ordinary talking, they must be grouped into syllables or words. A comparison of the sounds, syllables, or words recorded by the listener with those uttered by the speaker shows the fraction that is interpreted correctly. This fraction is called the articulation. It is syllable articulation S if the syllable is considered the unit—for example, "pat"—and sound articulation s if each individual speech sound is considered the unit—for example, "*p*," "*a*," and "*t*." In a similar way we may deal with the vowel articulation v or with the consonant articulation c. Various types of syllable lists have been used by various investigators.[5] A sample list of syllables of the consonant-vowel-consonant type is shown in Fig. 1. The carrier sentence shown in the second column is spoken before the "called" (i.e., spoken) syllable shown in the fourth column. In this sample test there were 6 vowel errors out of 66 spoken, 35 consonant errors out of 132, 41 speech sound errors out of 198, and 32 syllable errors out of 66. So the values of S, s, v, and c are those shown in Fig. 1.

In this paper there are considered four extensive sets of articulation data designated I, II, III, and IV. The first three sets were taken at the Bell Telephone Laboratories, I in the years 1919 to 1925, II in the years 1928 to 1929, and III in the years 1935 to 1937. The IV set was taken at the Psycho-Acoustic Laboratory at Harvard University during World War II.

To obtain a desirable precision[5] in the measurement of articulation, it is advisable to use at least five different voices and five listeners, at least 25 values being averaged in some way to obtain a final value for the condition tested. Various types of weighting have been proposed and used but in this paper only straight averages without weighting are used.

In set I of the articulation tests (conducted by J. B. Kelly) each list was composed of 80 syllables of the consonant-vowel-consonant type, 10 syllables of the consonant-vowel type, and 10 syllables of the vowel-consonant type. For these tests the syllable and sound articulations are designated S_{23} and s_{23}, respectively, the subscript indicating that both two-sound syllables and three-sound syllables were used. The systems tested are designated I–A, I–B, I–C, etc., where A, B, or C represents added digits or letters or combination of these to differentiate the various systems that were tested at this time. In each of these tests only two talkers were used, a man and a woman, or in some instances two men, and seven or eight listeners. In the set II tests (conducted by W. B. Snow and A. Meyer) the lists were all of the consonant-vowel-consonant type. For these tests the syllable and sound articulations are designated S_3 and s_3, respectively, and the systems tested are designated II–A, II–B, II–C, etc. In these tests the crew of eight persons provided eight talkers, four men and four women, and eight listeners of whom four listened to each talker. Each articulation test thus yielded a value which was the average of 32 talker-listener pairs. In set III the lists were the same as for set II but a machine[6] was used for recording and calculating the average sound articulation, which is designated s_{3M}. In all the other tests the syllables heard were written by the listeners but in set III the listeners did not write but punched keys. This difference in technique is important as it yields slightly different values, and so the subscript M is added to s_3. No values of syllable articulation S_{3M} were recorded because the machine was not so arranged. The systems tested in set III are designated III–A, III–B, III–C, etc. In tests III the crew consisted of from six to eight talkers and six to eight listeners, with 32 talker-listener pairs per test as in tests II. In set IV, the tests made

[5] H. Fletcher and J. C. Steinberg, "Articulation testing methods," Bell Sys. Tech. J. 8, 806 (1929).

[6] T. G. Castner and C. W. Carter, "Developments in the application of articulation testing," Bell Sys. Tech. J. 12, 347 (1933); L. Y. Lacy, "Automatic articulation testing apparatus," Bell Lab. Record 12, 276 (1934).

at Harvard University,[4] the lists were composed of consonant-vowel and vowel-consonant syllables. For these tests the syllable and sound articulations are here designated S_2 and s_2, respectively. The systems tested are designated H–A, H–B, H–C, etc., the H indicating that the tests were made at Harvard University. In these tests two male talkers and six male listeners were used. Table I is a summary of these groups of articulation tests, together with the dates of the tests and the designations of the various articulation values.

2. THE RELATION BETWEEN ARTICULATION INDEX AND THE VARIOUS ARTICULATION VALUES

These various articulation values may be considered as probabilities. For example, the sound articulation s is the probability that a speech sound will be interpreted correctly and the syllable articulation S is the probability that a syllable will be interpreted correctly. Similarly the consonant articulation c and the vowel articulation v are, respectively, the probability that a consonant and the probability that a vowel will be interpreted correctly. If a syllable is composed of one consonant and one vowel, the probability that the syllable will be interpreted correctly is equal to the product cv.

From certain standpoints, the sound error $e=1-s$ may be regarded as a probability. This will be understood from the following considerations. As is well known, the speech sounds are transmitted from speaker to listener by waves having frequency components

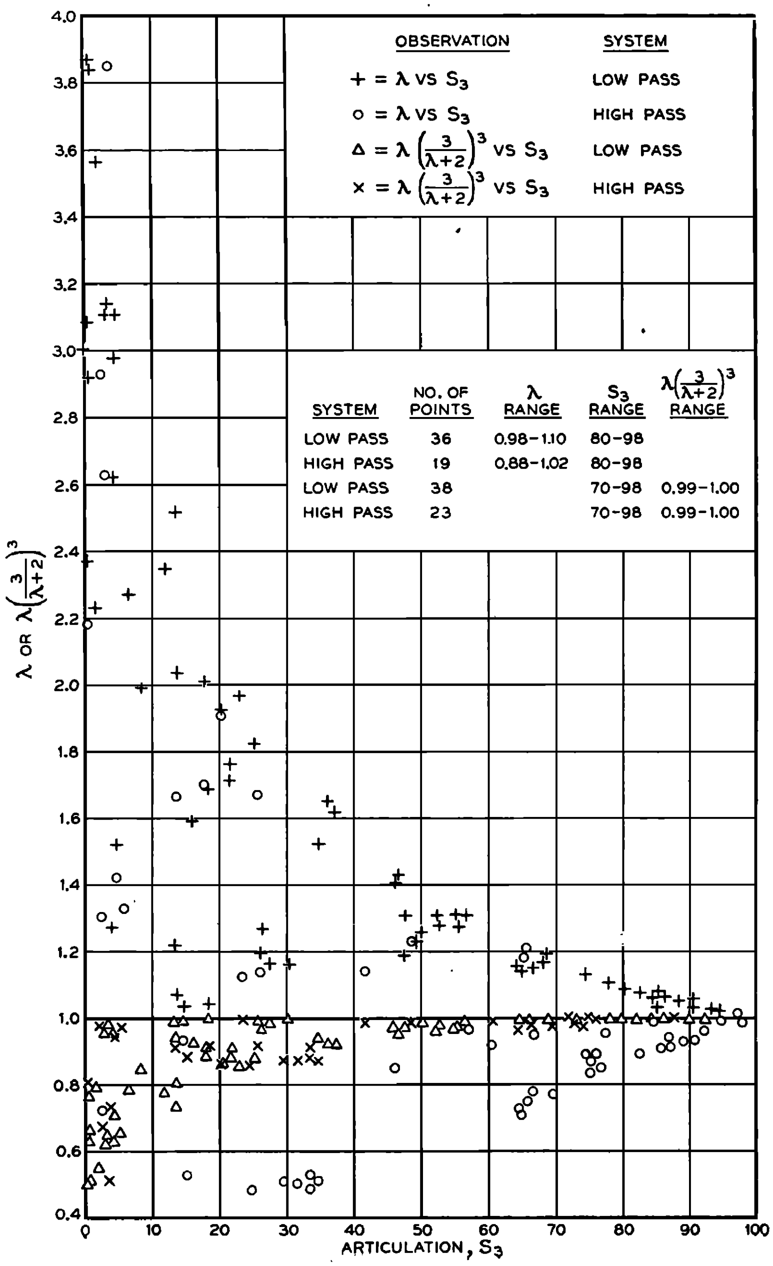


FIG. 2. Observed values of λ and $\lambda(3/\lambda+2)^3$ *vs.* syllable articulation S_3 for filter systems.

TABLE II. Maximum values of s_{3M} for ideal filter systems.

Cut-off frequency estimated from response curve	Estimated correction Δf	Ideal filter cut-off frequency f_c	Observed sound articulation s_{3M}
	High pass filter systems		
520 c.p.s.	45 c.p.s.	565 c.p.s.	0.975
1000	25	1025	0.95
1410	15	1425	0.91
1410	15	1425	0.925
2510	10	2520	0.74
2510	10	2520	0.77
System III		285	0.981
	Low pass filter systems		
1040	0	1040	0.68
1040	0	1040	0.705
1920	15	1905	0.875
1920	15	1905	0.875
2930	45	2885	0.94
4550	150	4400	0.973
System III		6500	0.981

from about 100 c.p.s. to 8000 c.p.s. Consider the frequency range divided into n frequency bands. It has been found experimentally that if e_1 is the sound articulation error when only the first band is used, e_2 when only the second band is used, and e_k when only the kth band is used, then when all bands are used simultaneously, the sound articulation error e is given by the following product:

$$e = e_1 e_2 \cdots e_k \cdots e_n. \tag{1}$$

In other words, e_1, $e_2 \cdots e_k \cdots e_n$ act as probabilities. This relationship was deduced by J. Q. Stewart in 1921 from articulation data on band pass filters which could not be considered as very accurate but the concept proved to be a useful one. It will be shown that only small corrections to this fundamental equation are necessary to fit the various kinds of articulation tests. For example, if 10 errors out of 100 spoken sounds are made when only band 1 is used, and 20 errors are made when only band 2 is used, then when both bands 1 and 2 are used simultaneously, the error is $e = 0.1 \times 0.2 = 0.02$, or two errors will be made.

As before, let s = sound articulation corresponding to error e. Then $e = 1 - s$; $e_1 = 1 - s_1$; etc. Hence Eq. (1) can be written thus:

$$1 - s = (1 - s_1)(1 - s_2) \cdots (1 - s_k) \cdots (1 - s_n)$$

from which the following relation is obtained:

$$\log(1 - s) = \sum_1^n \log(1 - s_k). \tag{2}$$

If we regard $\log(1 - s_k)$ as a property of the kth frequency band which measures the contribution of this band, then Eq. (2) states that the total contribution of the entire frequency band transmitted is equal to the sum of the contributions of the various partial frequency bands used singly. The term articulation index A has been used to designate such a simply additive measure of the contribution of a frequency band, and it is related to the quantity $\log(1 - s)$ by a factor of proportionality, as shown by the equation

$$A = -(Q/p)\log_{10}(1 - s). \tag{3}$$

The quantity $\log_{10}(1 - s)$ is always negative when $s > 0$; and the two quantities Q and p which enter into the factor of proportionality are each taken as positive in sign. Thus a negative sign before the right-hand member of Eq. (3) insures that the articulation index A will be positive in sign (if not zero).

The quantity Q in Eq. (3) is a constant and will now be evaluated.

The quantity p is called the proficiency factor for it is dependent upon the proficiency of the particular talker-listener pair combination. It is sometimes called the practice factor. It may be divided into two factors—namely, p_T, the proficiency factor for the talker; and p_L, the proficiency factor for the listener. Then

$$p = p_T p_L, \tag{4}$$

so if the talker speaks in a language unfamiliar to the listener $p_T = 0$ and so $p = 0$, and consequently s will be zero regardless of the value of the articulation index A. It is useful to divide p into these two factors when one desires to rate the proficiency of the talker and listener separately. For example, children in schools for the

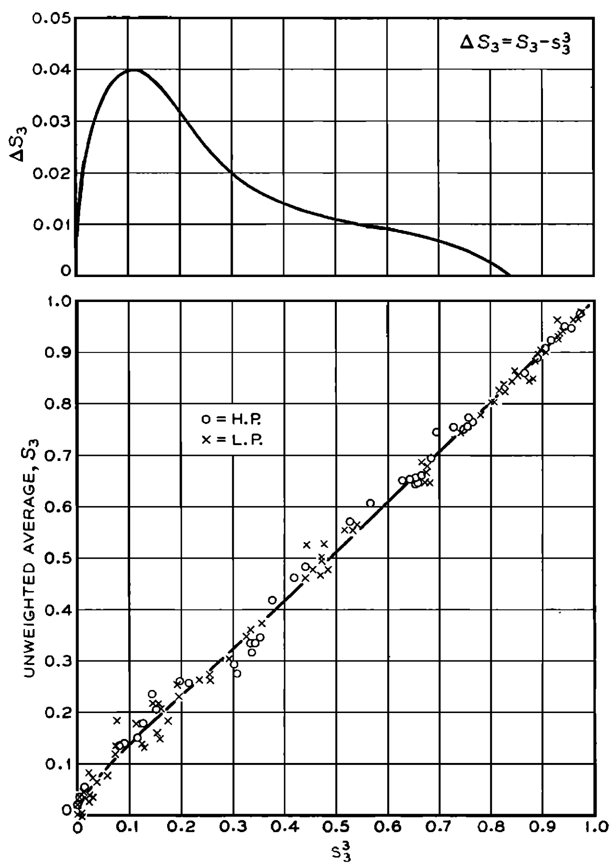
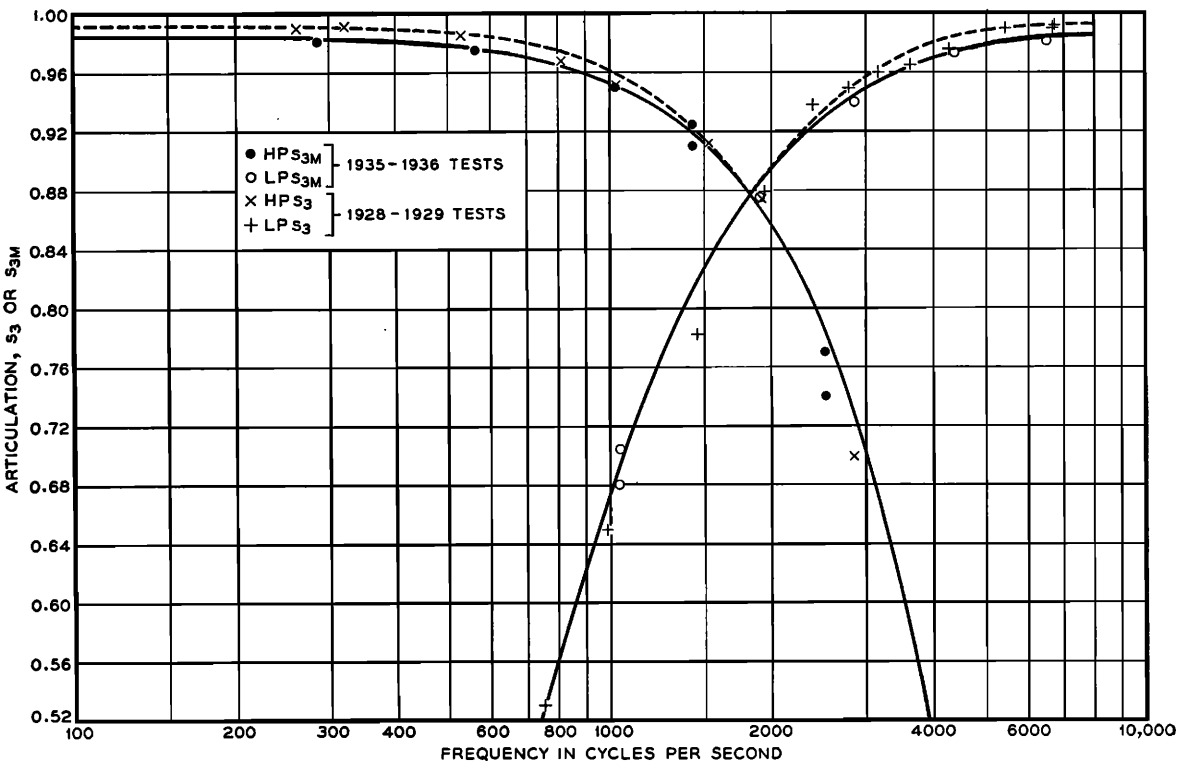


FIG. 3. Unweighted average syllable articulation S_3 and ΔS_3 *vs.* cube of speech sound articulation s_3^3 for filter systems.



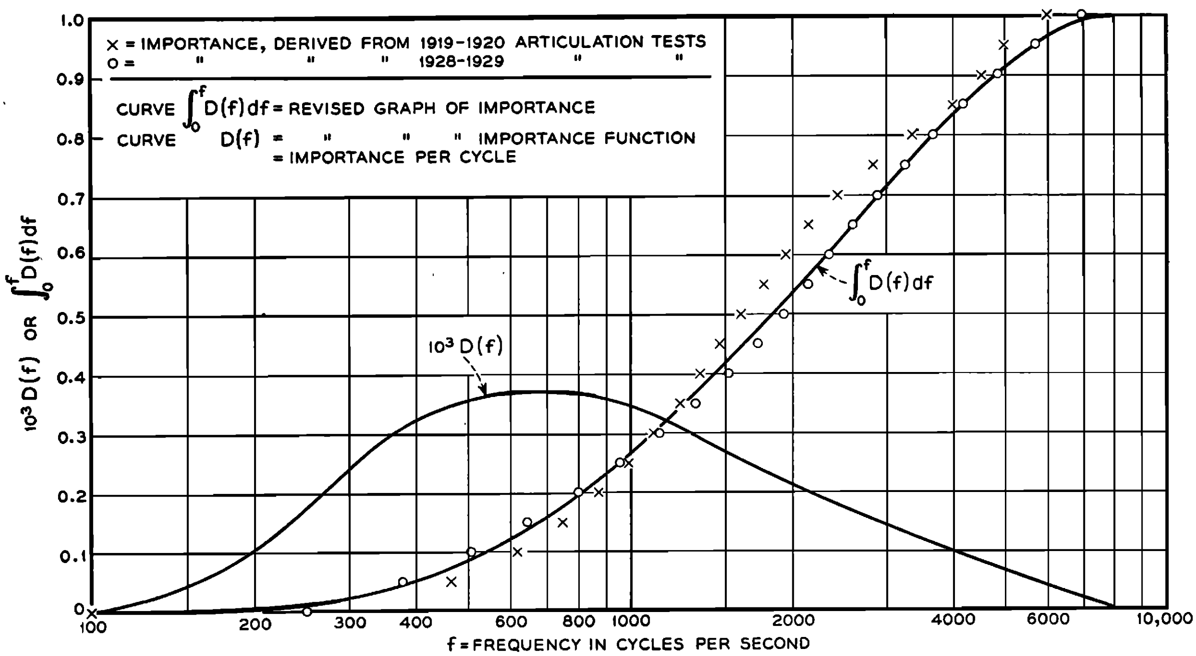


FIG. 4 (above). Speech sound articulation s_3 and s_{3M} *vs.* ideal filter cut-off frequency.

FIG. 5 (below). The importance of each frequency region for articulation or intelligibility. (See also Table XII.)

deaf can be rated on this scale. However, for calculating the articulation of telephone systems it is only necessary to know p. For a well-practiced and intelligent talker-listener pair the value p is taken arbitrarily equal to unity and Q is chosen so this is possible. Also the value of A for a flat response system having optimum amplification (that is, best for interpreting the speech sounds) is also taken arbitrarily equal to unity.*

* Although the present analysis of articulation phenomena assumes that the optimum telephone system is one having a uniform orthotelephonic response at all frequencies, other methods of analysis may make different assumptions. However, the writers

It has been found experimentally that when such a pair with $p=1.0$ make articulation tests over a system with $A=1.0$, then s is found to be 0.985. Consequently, from Eq. (3) the value of Q is found to be 0.55.

The 1935–1936 articulation crew of talkers and listeners was chosen as the crew for which the proficiency factor p shall be equal to unity. The average value of s for this crew corresponds to 0.985 for the flat response system at optimum amplification. Such a system will be called an ideal system and such a talker-listener pair a reference pair. For all other systems the value of A is between unity and zero and is a quantitative measure of the merit of the system for transmitting the speech sounds.

So when a talker uses a constant talking intensity level and a fixed transmission system transmits the speech sounds to a listener, then the articulation index A is constant and the two quantities, p the proficiency, and s the sound articulation obtained, vary together in accordance with Eq. (3). For a system with adjustable gain the maximum articulation index depends only upon the system and is independent of talkers and listeners. In a paper[7] (not yet published) delivered before the Acoustical Society of America in 1945, W. A. Munson showed that an equation similar to Eq. (3) can be deduced from the statistical properties of speech and hearing.

As will be shown later, Eq. (3) fits the experimental data taken for systems III. Thus the equation

$$s_{3M}=1-10^{-Ap/0.55} \tag{5}$$

holds without modification for data taken on systems III. For the results of articulation tests on systems other than III, slight modifications of Eq. (5) are necessary. Before considering the experimental data which show what these modifications must be, it is necessary to consider the probability relationships between the various types of articulation.

Syllables of the type consonant-vowel-consonant were also used for systems II but the sounds heard were recorded by writing rather than by punching a key so that the listener had a chance to change some of the letters after they had been written if he thought an error had been made. One would expect the techniques used in testing systems II and III would yield approximately the same results. They do except for high values where the data show s_3 to be slightly greater than s_{3M} but definitely more than the observational error. The values of s_{23} are also slightly higher than s_{3M}.

Consider the probability relationships existing between the various articulation values. In a syllable of the type consonant-vowel-consonant the syllable articulation S_3 is related to the consonant articulation c and the vowel articulation v by the equation

$$S_3=cvc. \tag{6}$$

Since there are twice as many consonants as vowels in such lists,

$$s_3=(2c+v)/3. \tag{7}$$

If we define λ by the equation

$$\lambda=v/c \tag{8}$$

then

$$S_3=\lambda(3/\lambda+2)^3 s_3^3. \tag{9}$$

Similarly for the Harvard lists, which were of the type consonant-vowel and vowel-consonant, the value of S_2 is given by

$$S_2=cv=\lambda(2/\lambda+1)^2 s_2^2. \tag{10}$$

For the lists used in set I there were 80 of the consonant-vowel-consonant type and 10 of the vowel-consonant type and 10 of the consonant-vowel type so

$$S_{23}=\tfrac{1}{5}cv+\tfrac{4}{5}cvc \tag{11}$$

and

$$S_{23}=\tfrac{1}{5}\lambda(14/5\lambda+9)^2 s_{23}^2+\tfrac{4}{5}\lambda(14/5\lambda+9)^3 s_{23}^3. \tag{12}$$

All of these probability relations are dependent upon the proposition that the probability of interpreting one sound is independent of the other sounds in the syllable. However, the data indicate that this is not strictly true so corrections for this lack of independence must be made.

These relations are dependent upon λ, which is known to vary through wide limits. For example, in Fig. 2 are shown all the data taken on filter systems II. The values of S_3 are shown as abscissas and the values of λ and $\lambda(3/\lambda+2)^3$ as ordinates. The values of λ are scattered but most of the points for the high pass filter systems are below unity while those for the low pass filter systems are above unity. These varying values of λ make it seem hopeless to obtain unique relations between s_{3M}, S_3, s_3, S_2, s_{23}, and S_{23}. However, the case is not as bad as it looks for the factor $\lambda(3/\lambda+2)^3$ varies only from unity to 0.9 as λ varies from 0.6 to 1.75. It is seen to be unity for values of S_3 greater than 0.6. However, there are other influences—one referred to above being the lack of independence of one sound articulation upon the other two in the syllable—which have a greater effect than the effect of the departure of λ from unity. That this is true is shown by the curve in Fig. 3, which shows a plot of S_3 *versus* s_3^3 for all the set II filter data. According to Eq. (9) S_3 should always be less than s_3^3 but the curve through the data shows that S_3 is always either greater than or equal to s_3^3. Consequently one must rely upon the experimental data rather than these statistical relations but the latter are very useful because they hold except for small corrections which will now be considered. There is an influence, which we will call influence X, tending to make S_3 larger than s_3^3, which more than overcomes the

know of no articulation observations which are sufficiently accurate to distinguish between these assumptions.

[7] W. A. Munson, "Relation between the theory of hearing and the interpretation of speech sounds," J. Acous. Soc. Am. 17, 103A (1945).

effect of λ. It also affects the relation between s_{23} and S_{23} and s_2 and S_2.

Equation (5) is confirmed by experiments with filter systems from the series III tests, as now will be shown. An experimental curve of s_{3M} *versus* gain was obtained for each filter system and the maximum value of s_{3M} (corresponding to the optimum gain for interpretation) was determined from these curves. The values are given in Table II.

Since the response of the filter systems only approached the ideal, certain small corrections to the cut-off frequencies were made as indicated in Table II in order to reduce the results to those which would be obtained if they were ideal. These corrections were first estimated from the lack of flatness of the response of the filter systems and the lack of steepness at the cut-off frequency. After the method of calculation here being developed was available, the estimates were made to agree with the calculated correction. These results are plotted in Fig. 4 by circles and dots and the two solid lines were considered to fit the points. The points corresponding to the +'s and ×'s will be discussed

TABLE III. Relationship between articulation and articulation index.

$s_{3M}=1-10^{-Ap/0.55}$

$s_3=s_{3M}+\Delta s_3$, $s_{23}=s_{3M}+\Delta s_{23}$, $s_2=s_{3M}+\Delta s_2$

$S_3=s_3^3+\Delta S_3$, $S_{23}=0.8s_{23}^3+0.2s_{23}^2$, $S_2=s_2^2$

1 Ap	2 s_{3M}	3 Δs_3	4 s_3	5 ΔS_3	6 S_3	7 Δs_{23}	8 s_{23}	9 S_{23}	10 Δs_2	11 s_2	12 S_2	13 I
1.10	0.990	0.006	0.996	0	0.988	0.006	0.996	0.989	0	0.990	0.980	0.999
1.05	0.988	0.006	0.994	0	0.982	0.006	0.994	0.983	0	0.988	0.976	0.999
1.00	0.985	0.007	0.992	0	0.976	0.007	0.992	0.978	0	0.985	0.970	0.999
0.98	0.983	0.008	0.991	0	0.973	0.008	0.991	0.975	0	0.983	0.967	0.999
0.96	0.982	0.008	0.990	0	0.970	0.008	0.990	0.972	0	0.982	0.964	0.998
0.94	0.980	0.009	0.989	0	0.967	0.009	0.989	0.969	0	0.980	0.961	0.998
0.92	0.979	0.009	0.988	0	0.964	0.009	0.988	0.966	0	0.979	0.958	0.997
0.90	0.977	0.010	0.987	0	0.961	0.010	0.987	0.963	0	0.977	0.955	0.997
0.88	0.875	0.010	0.985	0	0.956	0.010	0.985	0.959	0.001	0.976	0.952	0.996
0.86	0.973	0.010	0.983	0	0.950	0.010	0.983	0.953	0.001	0.974	0.949	0.996
0.84	0.970	0.011	0.981	0	0.943	0.011	0.981	0.947	0.002	0.972	0.945	0.996
0.82	0.968	0.010	0.978	0	0.935	0.010	0.978	0.939	0.002	0.970	0.941	0.995
0.80	0.965	0.010	0.975	0	0.926	0.010	0.975	0.931	0.003	0.968	0.937	0.995
0.78	0.962	0.009	0.971	0	0.916	0.009	0.971	0.921	0.004	0.966	0.932	0.994
0.76	0.958	0.009	0.967	0	0.906	0.009	0.967	0.910	0.004	0.962	0.926	0.994
0.74	0.955	0.008	0.963	0	0.894	0.008	0.963	0.900	0.004	0.959	0.920	0.993
0.72	0.951	0.008	0.959	0	0.882	0.008	0.959	0.890	0.005	0.956	0.914	0.993
0.70	0.947	0.007	0.954	0	0.868	0.007	0.954	0.877	0.005	0.952	0.906	0.992
0.68	0.942	0.007	0.949	0	0.854	0.007	0.949	0.864	0.006	0.948	0.898	0.992
0.66	0.937	0.006	0.943	0	0.839	0.007	0.944	0.850	0.006	0.943	0.889	0.991
0.64	0.931	0.006	0.937	0.001	0.824	0.007	0.938	0.836	0.007	0.938	0.879	0.991
0.62	0.925	0.005	0.930	0.002	0.806	0.007	0.932	0.822	0.007	0.932	0.869	0.990
0.60	0.919	0.003	0.922	0.003	0.787	0.007	0.926	0.807	0.007	0.926	0.858	0.990
0.58	0.912	0.002	0.914	0.004	0.767	0.007	0.919	0.790	0.008	0.920	0.846	0.988
0.56	0.904	0.001	0.905	0.005	0.746	0.007	0.911	0.771	0.008	0.912	0.832	0.986
0.54	0.896	0	0.896	0.006	0.725	0.007	0.903	0.752	0.008	0.904	0.818	0.984
0.52	0.887	0	0.887	0.007	0.704	0.007	0.894	0.732	0.009	0.896	0.802	0.982
0.50	0.877	0	0.877	0.007	0.681	0.008	0.885	0.711	0.009	0.886	0.785	0.980
0.48	0.866	0	0.866	0.008	0.657	0.009	0.875	0.689	0.009	0.875	0.767	0.976
0.46	0.854	0	0.854	0.009	0.632	0.010	0.864	0.665	0.010	0.864	0.747	0.971
0.44	0.842	0	0.842	0.009	0.606	0.011	0.853	0.642	0.010	0.852	0.726	0.967
0.42	0.828	0	0.828	0.010	0.578	0.013	0.841	0.617	0.011	0.839	0.704	0.964
0.40	0.813	0	0.813	0.010	0.547	0.014	0.827	0.589	0.011	0.824	0.679	0.960
0.38	0.796	0	0.796	0.011	0.515	0.015	0.811	0.558	0.012	0.808	0.653	0.954
0.36	0.778	0	0.778	0.012	0.483	0.016	0.794	0.527	0.012	0.790	0.625	0.948
0.34	0.759	0	0.759	0.013	0.450	0.018	0.777	0.497	0.013	0.772	0.596	0.940
0.32	0.738	0	0.738	0.014	0.416	0.022	0.760	0.467	0.013	0.751	0.564	0.932
0.30	0.715	0	0.715	0.016	0.382	0.027	0.742	0.437	0.014	0.729	0.531	0.925
0.28	0.690	0	0.690	0.018	0.347	0.033	0.723	0.407	0.014	0.704	0.496	0.91
0.26	0.663	0	0.663	0.021	0.312	0.041	0.704	0.378	0.015	0.678	0.460	0.90
0.24	0.634	0	0.634	0.024	0.279	0.049	0.683	0.348	0.015	0.649	0.422	0.87
0.22	0.602	0	0.602	0.029	0.247	0.059	0.661	0.318	0.016	0.618	0.382	0.84
0.20	0.567	0	0.567	0.034	0.216	0.071	0.638	0.289	0.018	0.585	0.342	0.80
0.18	0.529	0	0.529	0.038	0.186	0.085	0.614	0.261	0.023	0.552	0.304	0.76
0.16	0.488	0	0.488	0.040	0.156	0.101	0.589	0.233	0.031	0.519	0.269	0.67
0.14	0.444	0	0.444	0.039	0.127	0.118	0.562	0.205	0.040	0.484	0.234	0.57
0.12	0.395	0	0.395	0.036	0.098	0.137	0.532	0.177	0.051	0.446	0.199	0.46
0.10	0.342	0	0.342	0.031	0.071	0.155	0.497	0.148	0.063	0.405	0.164	0.39
0.08	0.285	0	0.285	0.026	0.049	0.172	0.457	0.118	0.074	0.359	0.129	0.29
0.06	0.222	0	0.222	0.020	0.031	0.186	0.408	0.088	0.083	0.305	0.093	0.22
0.04	0.154	0	0.154	0.014	0.018	0.194	0.348	0.058	0.088	0.242	0.058	0.16
0.02	0.080	0	0.080	0.007	0.008	0.183	0.263	0.028	0.072	0.152	0.023	0.08
0.00	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00

TABLE IV. Maximum values of s_3 for ideal filter systems.

Cut-off frequency estimated from response curve	Estimated correction Δf	Ideal filter cut-off frequency f_c	Observed sound articulation s_3
	High pass filter systems		
250 c.p.s.	70 c.p.s.	320 c.p.s.	0.991
485	45	530	0.986
775	35	810	0.968
1000	30	1030	0.952
1500	25	1525	0.912
1900	15	1915	0.874
2850	15	2865	0.699
System II		260	0.990
	Low pass filter systems		
755	0	755	0.53
1000	10	990	0.65
1480	20	1460	0.782
1980	30	1950	0.879
2440	40	2400	0.938
2850	50	2800	0.949
3250	65	3185	0.960
3700	75	3625	0.965
4400	100	4300	0.975
5600	160	5440	0.989
7000	250	6750	0.991
System II		6750	0.990

later. From this pair of solid-line curves one can obtain the function D which expresses at each frequency the importance of that frequency region for articulation.

This function is defined by the equation

$$D=dA_f/df, \tag{13}$$

where A_f is the maximum articulation index of an ideal filter system and f the frequency of cut-off. In other words, dA_f is the amount of articulation index carried by the small frequency band df in the frequency region between f and $f+df$ when the speech band is delivered to the ear at the optimum level for interpretation. For a low pass filter system with cut-off frequency f, then

$$A_f=\int_0^f Ddf. \tag{14}$$

From Eqs. (13) and (5) the expression for the function D becomes

$$D=(1/p)(0.239/1-s_{3M})(ds_{3M}/df). \tag{15}$$

The values of s_{3M} corresponding to $p=1.0$ are shown by the solid-line curves in Fig. 4. When the same curves are replotted using a linear scale of frequencies, the slopes of the curves give ds_{3M}/df, and so the function D is determined at each frequency. A plot of the values of D is shown in Fig. 5 by the curve $10^3D(f)$ *versus* f. Also, the values of A_f from Eq. (14) are shown on this plot by the integral curve. Before drawing this final integral curve, three sets of data of the type shown in Fig. 4 were examined and the curve in Fig. 5 was chosen as giving the best fit for all of the facts known. The points indicated by circles in Fig. 5 define the earlier curve derived by Steinberg, Galt, and Anderson in 1937 and published in a paper by French and Steinberg.[3] The points indicated by crosses define a still earlier curve obtained by Fletcher in 1921 and given in an unpublished memorandum.

Having the importance function D determined, the pair of curves for s_{3M} in Fig. 4 now can be computed from Eq. (5), with $p=1.0$ so that

$$s_{3M}=1-10^{-A_f/0.55} \tag{16}$$

where A_f at each frequency is given by the integral curve in Fig. 5. The solid lines in Fig. 4 are such calculated curves. The agreement between the calculated curves and the observed points justifies the use of these curves in finding the derivative ds_{3M}/df from which the importance function D was obtained. Thus it is seen that Eq. (16) fits the set III data.

It will be realized that when the frequency importance function D has been integrated over a chosen frequency region from f_1 to f_2, the integral $\int_{f_1}^{f_2}Ddf$ is the same as the articulation index function, if the system is without noise and is composed of distortionless elements, and if it has an ideal uniform response at all frequencies with the over-all gain adjustment corresponding to maximum articulation for the ideal system. Consider a frequency such that the articulation of an ideal high pass filter system having this cut-off frequency is just equal to the articulation of an ideal low pass filter system having the same cut-off. These two systems having equal articulations must have equal values of the articulation index, each value being 0.5. The cut-off frequency of this pair of filters should agree with the intersection of the two curves in Fig. 4. The ordinate of the point of intersection is 0.876. When this value is substituted for s_{3M} in Eq. (16), the corresponding value of A_f is found to be 0.5 as it should be.

Because of the inherent lack of precision of articulation tests which form the experimental basis for the frequency importance function D, it is a matter of interest to discuss further the derivation of the function from this standpoint. Whether the function is derived by the method used in this paper or by the earlier procedure used in 1921 and also in 1937 and described by French and Steinberg,[3] the basic data are of the same type and the first treatment of the data is the same. A series of articulation tests, over a wide range of gain adjustments, is made upon a high quality telephone system containing a filter of known cut-off frequency, and the results are shown as discrete points in a plot of articulation *versus* gain. Similar plots are made for various other cut-off frequencies. In any plot, each point shows the average of a certain number of values of articulation observed by each talker-listener combination. For example, in a plot of the set II tests of filtered systems each point shows the average of thirty-two talker-listener observations of the syllable articulation S_3. For points near the maximum articulation of a filter system the typical r.m.s. deviation σ of one observation from the average is about 0.02 to 0.05 when the suppressed frequency region is narrow, and about 0.07 to 0.15 when the suppressed region is wide. The values of $(1/n)^{\frac{1}{2}}\sigma$ for these points are typically from 0.005 to 0.025. These numbers are in terms of S_3, the maximum possible S_3 being 1.000. Through such an array of points a smooth curve is drawn, and the maximum ordinate of this curve is regarded as the maximum articulation for the cut-off frequency of the filter under test. Plots containing such arrays of points are shown in Figs. 25 and 27 for the set II filters, and in Figs. 24 and 26 for the set III filters. It must be remembered that the curves in these figures are not

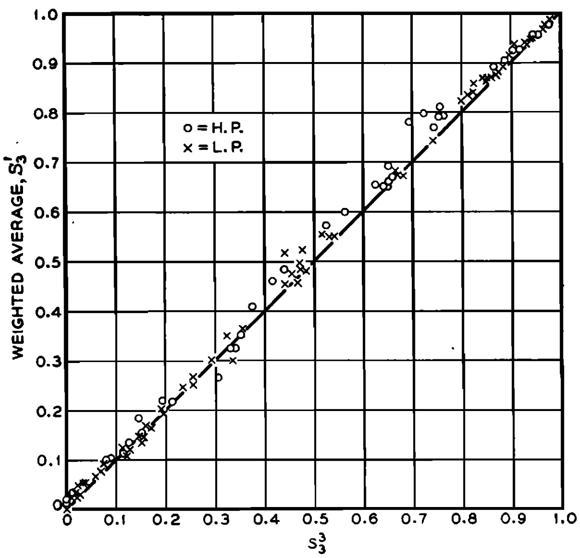


FIG. 6. Weighted average syllable articulation S_3' *vs.* cube of speech sound articulation s_3^3 for filter systems.

the curves just mentioned as having been drawn to represent the observed points. On the contrary, the curves in these figures were calculated using the functions adopted in the present paper, and hence their agreement or disagreement with the data in the regions of greatest articulation shows just how closely the curves in Fig. 5 represent the observed maxima. Although in each of these figures the response of the unfiltered system is included, the observed and calculated articulations are omitted to avoid crowding but can be found by comparison with Figs. 15 and 16.

When a smooth curve has been drawn to represent the array of points from articulation tests upon a filter with different adjustments of gain, it might be expected that the maximum ordinate of the curve would be determined with an uncertainty comparable in magnitude with the quantity $(1/n)^{\frac{1}{2}}\sigma$. In partial corroboration of this, it is found that repetitions of the set II filter tests at virtually the same gain, in the region of maximum articulation, generally yield values of S_3 which differ from the earlier values by an amount that does not exceed $(1/n)^{\frac{1}{2}}\sigma$. A few notable exceptions to this statement are found, which have raised the question whether or not in some instances other factors may have been operating,—for example, some general gain or loss of proficiency on the part of the crew members.

The derivation of the importance function by the method used in this paper proceeds to assemble the maximum values found for the various filters, forming a succession of discrete points in a plot such as the points for the set III filters in Fig. 4. A curve is drawn to represent the points derived from low pass filters and another curve for high pass filters, the two curves being not independent but instead necessarily related because each represents the same function. When both curves cannot be made to fit their respective sets of points, the fit of one curve or the other is sacrificed in the region of least precision. Thus the integral function in Fig. 5 is made to depend more upon high pass filter observations in the low frequency region, and more upon low pass filter observations at the higher frequencies.

In order to derive a single integral curve from the results of three different sets of filter system articulation tests, there must be some basis for weighting the data. The curve shown in Fig. 5 gives greatest weight to the set III data and least to the set I data, for the following reasons. The number of voices used in these groups of tests was:

Set	Voices per test	Total voices
I	2	3
II	8	12
III	8	9

Thus the set I data were of limited relative value from this standpoint. Moreover, the responses of the set I systems were less well known, and there were fewer tests from which the maximum could be determined in each plot of articulation *versus* gain. The talking levels and circuit conditions in the set III tests were better controlled than in either of the other sets of tests, and the responses were better known. The number of different gain adjustments used for each filter in set III was in general greater than in set II and much greater than in set I. Thus the points in Fig. 4 representing maximum articulation values for the set III filter

TABLE V. Relation between syllable and sound articulations for systems I.

S_{23}	s_{23} HQ	HP	LP	BP	Average	s_{23}^2	s_{23}^3	$0.2s_{23}^2 + 0.8s_{23}^3$
0.02				0.23	0.23	0.053	0.013	0.021
0.05				0.32	0.32	0.102	0.033	0.047
0.1	0.43			0.415	0.422	0.178	0.075	0.096
0.15	0.513		0.51	0.488	0.514	0.264	0.136	0.162
0.2	0.571		0.547	0.55	0.556	0.309	0.172	0.199
0.3	0.660		0.638	0.642	0.647	0.419	0.271	0.300
0.4	0.73		0.712	0.717	0.720	0.518	0.373	0.402
0.5	0.786		0.774	0.778	0.779	0.607	0.473	0.500
0.6	0.837	0.82	0.828	0.831	0.829	0.687	0.570	0.594
0.7	0.881	0.878	0.878	0.878	0.879	0.773	0.679	0.698
0.8	0.924	0.922	0.922	0.922	0.923	0.852	0.786	0.799
0.9	0.963	0.964	0.962	0.963	0.963	0.927	0.893	0.900
0.95	0.981	0.981	0.981	0.981	0.981	0.967	0.944	0.948

systems are regarded as more reliable than the corresponding maxima from set II and much more reliable than the maxima from set I (not shown in Fig. 4).

Having in mind these uncertainties associated with the data, an attempt has been made to re-draw the curves in Fig. 4 in such a way as to shift the intersection point as far to the right or to the left as the data would conceivably permit. This study led to the conclusion that on the basis of the available articulation observations the intersection frequency in Fig. 4, and therefore the frequency in Fig. 5 at which $\int_0^f Ddf=0.5$, could not justifiably be shifted either way by as much as 100 c.p.s., and probably not by as much as 75 c.p.s. At higher frequencies the uncertainty in c.p.s. may be as great; at lower frequencies the uncertainty seems to be less than 50 c.p.s.

Various writers have mentioned the resemblance between the curve $\int_0^f Ddf$ *versus* f (which involves the properties of both speech and hearing) and several other functions all of which are derived from tests of hearing or from a study of the ear. The resemblances that exist among these other functions (which involve only the ear, and do not depend upon speech) have been known for some time. The relationship between the articulation index, or the frequency importance function for articulation, on the one hand, and the various functions of hearing, on the other hand, was recorded in 1939 in an unpublished memorandum by R. H. Galt. A general conclusion regarding the relationship between the importance of different frequency regions for articulation and the distribution of nerve endings on the basilar membrane was stated in Appendix 2 of OSRD Report No. 502, dated March 26, 1942, submitted by Steinberg, Galt, and others. Attention was called to certain of these relationships by French and Steinberg in their paper[3] delivered in 1946 and published in 1947. A discussion of these relationships and their implications was given by R. H. Galt in an unpublished paper[8] delivered before the Acoustical Society of America.

The first relation between articulation and articulation index is that given by Eq. (5). The values of Ap and s_{3M} obtained by this equation are tabulated in Table III in columns 1 and 2. Although the value of A never exceeds unity, the value of Ap may be higher for very expert talker-listener combinations. Therefore values of Ap from 0 to 1.1 are tabulated. Data similar to those shown in Table II for s_{3M} were also available for s_3 and are shown in Table IV. A plot of these data is shown in Fig. 4 by the +'s and x's. The dashed curve was considered to fit the observed points. Below the intersection point the solid curve was considered to fit the points for s_{3M} and also those for s_3 within the experimental error.

The difference between corresponding ordinates of the curves s_3 and s_{3M} in Fig. 4 is designated Δs_3 and is given in Table III in column 3. Since s_3 cannot be greater than unity, Δs_3 must approach zero as s_3 and s_{3M} approach unity. The addition of the numbers in columns 2 and 3 gives the numbers in column 4, which are the values of s_3 in the terms of Ap in the first column.

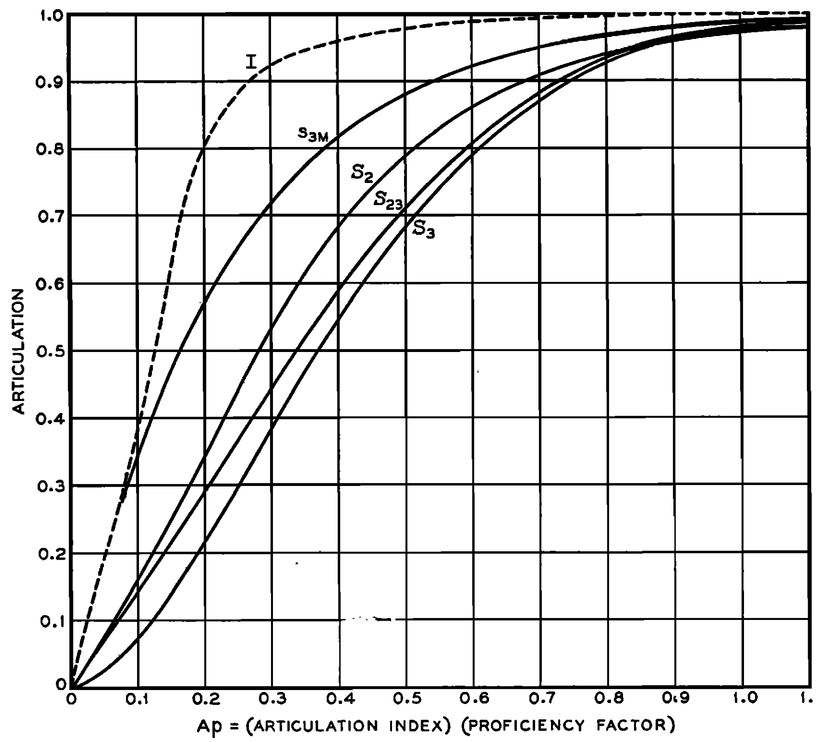


FIG. 7. Relation between various measures of articulation and the articulation index. The symbols for articulation are explained in Table I. The symbol I denotes sentence intelligibility.

[8] Rogers H. Galt, "The importance of different frequency regions for speech intelligibility," J. Acous. Soc. Am. 20, 592A (1948).

The relation between the syllable articulation S_3 and s_3^3 is obtained from the experimental data shown by discrete points in Fig. 3 which were obtained from articulation tests on a large variety of filter systems. The curve through the points was taken as fitting the experimental relation between S_3 and s_3^3. Above S_3 equal to 0.65 the value of S_3 will be seen to be accurately equal to s_3^3. Below this value S_3 is always greater than s_3^3 by an amount which will be designated ΔS_3. The values of ΔS_3 chosen are defined by the curve but it is evident from the scatter of the points that this is only an average. These values are given on an enlarged scale at the top of Fig. 3 and are recorded in column 5 of Table III. These values are added to the values of s_3^3 to give the values of S_3 recorded in column 6. It is important to notice that these relations are accurate for values of Ap above 0.5 but can be considered as only approximate for values below 0.5. Thus differences in this region between calculated and observed articulations will be expected to be of the same order as the scatter of the points in Fig. 3.

In order to take care of the skewness of the distribution of articulation values near zero and also near 100 percent, in the finding of an average of a group of values observed by different talker-listener pairs, a method of weighting these values was proposed in a paper[5] cited previously. Such weighted values were available for the systems II. In Fig. 6 a plot of the values of such weighted average $S_{3(\text{weighted average})}$ *versus* s_3^3 is shown. These weighted average articulations are taken from the same tests as the unweighted averages plotted in Fig. 3. A comparison between Figs. 3 and 6 indicates that $S_{3(\text{weighted average})}$ is more nearly equal to s_3^3 than is S_3 for the lower values, but for the higher values S_3 is closer to the values of s_3^3 than is $S_{3(\text{weighted average})}$. As stated above, all the values used in this paper are unweighted averages.

An examination of the corresponding values of S_{23} and s_{23} for systems I showed that Eq. (12) with λ equal to 1.0 held accurately over the entire range, or

$$S_{23}=0.2s_{23}^3+0.8s_{23}^3. \qquad (17)$$

The data in Table V confirm this relation. Two sets of data were taken upon the following four types of systems: (1) Approximately flat response system—HQ, (2) low pass filter systems—LP, (3) high pass filter systems—HP, and (4) band pass filter systems—BP. Curves showing the relation between s_{23} and S_{23} were drawn through all of the points for each of these types of systems. Values corresponding to each of the values of S_{23} shown in column 1 of Table V were taken from such sets of curves and are shown in columns 2–5, with the average for the four types of systems given in column 6. The values of S_{23} calculated from the values of s_{23} given in column 6 by Eq. (17) are given in the last column. A comparison of the first and last columns shows that formula (17) holds accurately not only for the region above $S_{23}=0.65$ but also for the entire range of values. (See also Fig. 125, p. 268 of reference 9). This is a surprising result in view of the relation between S_3 and s_3 in this region. The influence X referred to seems to just balance the effect of λ varying widely from unity.

Therefore, if we know the relation between s_{23} and s_{3M}, the values of s_{23} and S_{23} can be expressed in terms of Ap.

The values of s_{23} will be given in terms of $s_{3M}+\Delta s_{23}$. From the statistical relationship shown by Fig. 3, it would be expected that the values of Δs_{23} would be equal to Δs_3 in the region above $Ap=0.65$; they were so considered and are given in column 7 of Table III. The articulation data justify this assumption. In the lower range there are no experimental data giving a direct comparison between s_{23} and s_{3M} and thus giving Δs_{23} directly. So the values of Δs_{23} were chosen to give the best fit of the articulation data on a wide range of different systems. The values adopted are shown in column 7. These values of Δs_{23} are added to the values of s_{3M} to give the values of s_{23} given in column 8. The values of S_{23} are then calculated from these values of s_{23} by Eq. (17) and recorded in column 9.

In a similar way s_2 was obtained from s_{3M} by the addition of a correction term Δs_2. Since no values of s_2 were available to us we chose values of Δs_2 so that the articulation data given in terms of S_2 would best fit the calculated results. It is seen that these corrections are very small except for low values of s_2. In this region it is seen that Δs_2 is about $\frac{1}{2}\Delta s_{23}$. The values of Δs_2 are given in column 10. These values are added to s_{3M} to give the values of s_3 recorded in column 11. The values of S_2 are computed from these values of s_2 by the equation

$$S_2=s_2^2 \qquad (18)$$

and the values recorded in column 12.

Tests showing the relation between syllable articulation S_{23} and sentence intelligibility I have been made and recorded in the book on *Speech and Hearing*, page 266.[9] The value of I corresponding to S_{23} is given in the last column. It is the fraction of simple sentences (of the type used) which will be interpreted correctly.

So in Table III are contained the required relations between Ap, s_{3M}, s_3, S_3, s_{23}, S_{23}, s_2, S_2, and I. Several of these relations are shown by the curves of Fig. 7.

Before setting up methods of calculating A from the physical characteristics of the telephone system and noise conditions at the listener's ear, the methods of defining the response of the system, the speech intensity level of the talker, and the acuity level of the listener will be discussed, and how these together with the proficiency p were applied to the talkers and listeners and systems used in this investigation.

[9] H. Fletcher, *Speech and Hearing* (D. Van Nostrand Company, Inc., 1929).

3. THE RESPONSES OF THE TELEPHONE SYSTEMS

With the exception of three systems having carbon microphones, every articulation test system included here consisted of linear elements. Condenser microphones were used. The receivers were dynamic type earphones except in the 1919–1925 tests, when simple bipolar structures were used having special air damping.

The performance of any telephone system is here expressed in terms of its orthotelephonic response characteristic, which at each frequency is equal to the difference in db between the transmission supplied by the telephone system and that supplied by the orthotelephonic reference system.[3,10] This reference system consists of the air path between a talker and a listener, using one ear, who faces the talker in an otherwise free acoustic field at a distance of one meter between the lips of the talker and the aural axis of the listener.

In plotting each response *versus* frequency characteristic for a chosen adjustment of the amplifier gain, the adjustment has been designated arbitrarily as gain $\alpha=0$ db. The observed values of articulation have been plotted against the appropriate values of α.

In answer to the questions, how were these responses obtained? and how reliable are they?, a few remarks will be made regarding the three groups of telephone systems tested for articulation at the Bell Telephone Laboratories. For the fourth group, tested at Harvard University, reference should be made to publications by Egan and Wiener.[4]

The earliest of these groups of articulation tests was made in 1919–1920, when response determinations were in a relatively early stage of development. As compared with later tests, these responses are known with less certainty, especially at the higher frequencies. Over-all single frequency measurements (not published) were made by F. W. McKown in such a way as to compare the transmission of the telephone system with that of a one-half inch air path between an artificial voice and an artificial ear, the reference system at that time being the air path employed when a talker speaks from a distance of one-half inch directly into the ear of a listener.

As a check upon the measurements at single frequencies, an over-all comparison was made between the high quality system and the one-half inch air path using a talker and listener. Two voices were used and five listeners. The comparison test involved selecting the gain adjustment of the system that caused the two specimens of speech (which were of nearly identical quality) to sound equally loud. For five talker-listener pairs the gain for equal loudness of speech was lower than the gain based on single frequency tests by the following amounts:

$$1, 0, 3.5, 3.5, 2.5 \text{ db}; \quad \text{average}=2 \text{ db}$$

The single frequency responses have been used in the present study and have been converted to responses of the orthotelephonic type. The conversion was accomplished by adding two increments required by the change in the reference air path. The first increment takes account of the decrease in the acoustic pressure of received speech waves[11] under reference conditions when the distance measured from the lips to the ear is increased from one-half inch to one meter less the semi-diameter of the head. The second increment allows for the decrease in pressure at the ear caused when the listener turns through 90 degrees so as to face the speaker.[12,13] Finally, on the basis of limited and uncertain evidence, the single frequency response characteristics have been extended to the region above 4500 or 5000 c.p.s. as shown by broken lines in the attached plots of the 1919–1920 responses. For the 1919–1920 high quality system, here termed system I, the response is shown in Fig. 14.

The 1919–1920 high quality system was also used in the 1924–1925 articulation tests involving interfering pure tones, shown in Figs. 35–37.

For the two later groups (II and III) of articulation tests made at the Bell Telephone Laboratories, each over-all response may be regarded as obtained by adding together three responses measured separately, namely, the real voice response of the microphone, the electrical response of the circuit, and the real ear response of the receiver. Actually, the basic over-all response from the experimental standpoint was that of the 1935–1936 high quality system. From it the other over-all responses for this group are derived by adding to the high quality response the responses of the inserted networks together with small differences, if any, in the microphones and receivers as measured by coupler calibrations. Similarly, the over-all response of the 1928–1929 high quality system in Fig. 15 has been derived from that of the 1935–1936 system in Fig. 16 by adding coupler differences together with an increment (measured by C. W. Carter) which allows for a change of housing of the condenser microphone; other 1928–1929 over-all responses are then obtained by adding the responses of networks.

The experimental procedure followed in making real voice measurements of the 1935–1936 condenser microphone, and in making real ear measurements of the receiver, need not be given in detail because adequate accounts of the methods have been described in the literature on the subject.[3,4,14] In regard to the precision of the response measurements, it is presumed that the performance of the electrical circuits was known to within close limits, so that comments need be made

[10] A. H. Inglis, "Transmission features of the new telephone sets," Bell Sys. Tech. J. **17**, 374 (1938).

[11] H. K. Dunn and D. W. Farnsworth, "Exploration of pressure field around the human head during speech," J. Acous. Soc. Am. **10**, 184 (1939).

[12] L. J. Sivian and S. D. White, "On minimum audible sound fields," J. Acous. Soc. Am. **4**, 288 (1933).

[13] F. M. Wiener, "On the diffraction of a progressive sound wave by the human head," J. Acous. Soc. Am. **19**, 143 (1947).

[14] F. F. Romanow, "Methods of measuring the performance of hearing aids," J. Acous. Soc. Am. **13**, 294 (1942).

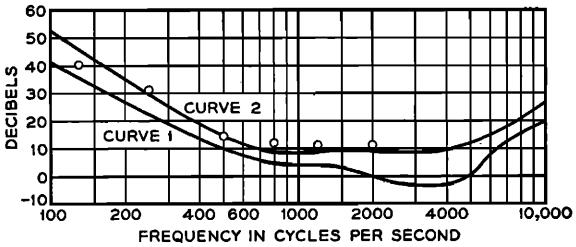


FIG. 8. Minimum audible pressure levels for typical listener of the articulation test crews, in decibels from 0.0002 dyne/cm². Curve 1 = Monaural minimum audible field. Curve 2 = Monaural minimum audible pressure level under earphone. Points = Observations for 1919–1920 crew.

only concerning the real voice and real ear measurements.

The over-all response given in Fig. 16 for the 1935–1936 high quality system is the mean of two independent determinations of this response. Each determination contained a measurement of the real voice response of the microphone and of the real ear response of the receiver, involving many differences in the technique of measurement and in the personnel of the testing crews.

The precision of a set of measurements of the real voice response of a microphone will be illustrated by the first determination used here, from data obtained by H. K. Dunn. Five voices were employed, and the spectrum below 8000 c.p.s. was divided into eleven octave or half-octave bands. The response was measured for every voice for each of the eleven bands. The r.m.s. deviation of the response for one voice from the mean of the responses for the five voices was obtained for each frequency band, and the average value of this deviation for the eleven bands was found to be 1.2 db, the maximum value for any band being 2.3 db and the minimum 0.4 db.

The precision of a set of measurements of the real ear response of a receiver will be illustrated for the first of the two determinations used here. This response was the average of four values of response—one value obtained by S. D. White using threshold observations made by a crew of eight listeners, and three values obtained by W. Koenig and H. J. Michael from three sets of loudness balances made by a different crew of seven listeners. At each frequency, the response obtained by one listener using two or more trials was regarded as one observation; thus there were eight (or seven) observations in each test, and the r.m.s. deviation of a single observation from the mean was found. These deviations are in general smaller for the intermediate frequencies from 600 c.p.s. to 2000 c.p.s. than for the extremes, and tend in the aggregate to be from 3 to 6 db, with a maximum of 8 db and a minimum of 1.1 db. At most frequencies the set of threshold observations scattered somewhat more widely than the best two of the three sets of loudness balances. The four mean values obtained from the four sets of data were averaged at each frequency to obtain the response here referred to as the first determination of the receiver response. Each group of four mean values thus averaged was in general spread over a range of 2 to 5 db.

The first determination of the over-all response of the high quality telephone system tested in 1935–1936 was obtained by adding the microphone and receiver responses, of which the precision has just been indicated, together with the response of the electric circuit. The second determination of the over-all response was the result of other measurements, one difference being that the real voice response of the microphone was measured by W. Koenig using bands of width 200 c.p.s. throughout the spectrum. Either of the two determinations generally differed from the final average by less than 1 db below 1000 c.p.s. and by less than 3db above 1000 c.p.s.

The responses of the 1936–1937 telephone systems in Figs. 21–23, 38, and 39 were derived from real voice (band width 200 c.p.s.) and real ear measurements made under the direction of W. Koenig and P. V. Dimock. Several units of each type of carbon microphone were tested, and the average or typical response was adopted. Each "hybrid" microphone (designated as HY in figures) consisted of the mechanical elements of the corresponding type of carbon microphone, but with the carbon granules either removed (transmitter types 323 and 625) or retained but not connected electrically (type 395), and including a small condenser microphone, so as to provide essentially a linear instrument which resembled the carbon instrument in geometry and motional characteristics.

The noise employed in certain of the tests represented in Figs. 22, 23, 38, and 39 was introduced electrically into the circuit from a phonograph record of typical room noise. The plotted spectrum levels are the levels of a field equivalent to the noise reaching the ear from the receiver.

4. ACUITY OF HEARING OF THE LISTENERS

The available information concerning the acuity of hearing of the listeners in the articulation tests will be summarized in the present section of this paper. The information is not sufficient to permit the assigning of different measures to the different crews, but is sufficient to justify the adopting of a definite acuity at each frequency, which acuity is regarded as characteristic of each of the testing crews.

The assumption that the average acuity of hearing was the same for each crew is made plausible by the fact that every crew member was a selected young adult, and that each crew contained from six to nine listeners so that the effects of small random individual peculiarities would tend to cancel out in the averages of observed articulations. In the 1919–1920 group of articulation tests, antedating the audiometer, the requirement to be satisfied by candidates in order that they might be accepted as listeners was that their performance must not be abnormal in a series of rehearsal tests upon the high quality system with different gain adjustments. The members of the 1928–1932 and 1935–1937 articulation crews were selected on the basis of having normal audiograms, with some attention given also to their interest in the work and to their alertness and general facility of performance in rehearsal tests.[5] At Harvard University the acuity of the observers was described as follows: "All of the crew members had satisfactorily normal hearing."[4] It is understood that the candidates for the Harvard articulation crew were accepted only if the audiograms were normal.

In order to assign a useful measure to the acuity of hearing of the typical listener in the articulation tests, it should be understood that for any person having a definite ability to hear, there may be widely different measures of acuity depending upon the technique and circumstances of making the measurements and upon the skill of the listener developed through experience in observing thresholds and in making other auditory observations. For the purposes of the present study, two different measures of acuity are required in order to answer the two following questions. First, at each frequency what was the minimum audible pressure as measured through the use of a refined technique in an extremely quiet place, the listener being experienced in observing thresholds and having his ears well rested against auditory fatigue? Second, at each frequency what was the hearing loss of the same listener measured by an audiometer with standard technique, the listener being inexperienced in observing thresholds and taking only ordinary precautions against noise and auditory fatigue? The first of these measures of acuity is significant in considering the effect upon articulation of a noise having a known intensity or masking spectrum. The second measure is significant in comparing the articulation observed by the test crews with that which would be observed by a different listener who is not skilled in observing thresholds but has an audiogram measured in the customary manner.

The first of the two expressions for the acuity of the

TABLE VI. Audiogram of typical articulation crew listener before training

(1) Frequency c.p.s.	(2) Audiogram hearing loss: Observed using 2A audiometer db	(3) Audiogram hearing loss: Proposed standard audiometer db	(4) Minimum audible pressure level under earphone: From column 2 per search tube calibration db	(5) Minimum audible pressure level under earphone: From curve 2 Fig. 8 db	(6) Difference Column 4–Column 5 db
125 or 128	−3.4	−4.9	44.1	46.7	−2.6
250 or 256	−1.0	−7.0	38.3	29.2	9.1
500 or 512	−0.9	−6.0	28.1	14.1	14.0
1000 or 1024	−5.0	−4.5	17.6	8.3	9.3
2000 or 2048	−4.0	−2.2	19.6	9.0	10.6
4000 or 4096	−0.7	−4.4	20.0	9.6	10.4
8000 or 8192	0.0	−3.1	25.9	21.2	4.7
Weighted average for speech	−4.0	−3.9			

TABLE VII. Talking levels and proficiency factors for various groups of articulation tests.

Date of tests	Acoustic talking level of crew measured or estimated	Proficiency factor p
1919–1920	*69 db from 10^{-16} watt/cm²	0.88 to 1.0
1924–1925	*68 to 70	0.95 to 1.05
1928–1929	69	1.0
1935–1936	68.5 to 69.5	1.0
1936–1937	66 to 67.5	1.0
1944–1945 (Harvard)	70	1.0

* = estimated.

typical articulation test listener has been based upon the binaural zero loudness contour[15] of the American Standards Association. This contour corresponds to the thresholds of observers who are somewhat more acute than the average young adult having normal hearing. From various threshold tests made by W. A. Munson, it was concluded by the present writers that in the frequency region more important for speech perception the binaural field threshold level for a typical experienced young listener is generally from 1.5 to 3 db above the zero loudness contour level. In the present work this difference has been arbitrarily adopted as 2.5 db at every frequency. Thus a curve was drawn parallel to the A.S.A. zero loudness contour but above it by 2.5 db to represent the free field intensity levels of pure tones at the threshold of audibility for two-ear listening, for the typical listener of the articulation test crews. From this curve for two ears, the corresponding free field intensity levels at threshold for one-ear listening have been found by adding the difference (1 ear−2 ears) given by French and Steinberg.[3]

The curve thus derived is plotted as curve 1 in Fig. 8. This curve shows the free field intensity level of a pure tone which is just audible to a typical member of the articulation test crews facing the source and listening with one ear, the intensity being measured with the observer absent. The technique for observing the thresholds indicated by this curve is a refined technique similar to the procedures followed by W. A. Munson in threshold observations connected with studies of loudness and masking. Two different procedures have been described. By one technique[16] the tone level is lowered near the threshold by successive steps of 1 db until a reversal occurs—that is, until the tone becomes inaudible. Then the tone level is raised by steps of 1 db until a second reversal occurs, the tone becoming audible. The entire procedure is repeated until the reversal points are definitely located. The average of the levels corresponding to the ascending and descending reversal points is accepted as the threshold intensity level. By the second technique[17] the

[15] "American standard for noise measurement," J. Acous. Soc. Am. 14, 109 (1942).
[16] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," J. Acous. Soc. Am. 5, 90 (1933).
[17] J. C. Steinberg and W. A. Munson, "Deviations in the loudness judgments of 100 people," J. Acous. Soc. Am. 8, 71 (1936).

observer watches a signal light and by pressing a button indicates whether or not he hears a tone which is presented for a duration time of 1 second while the light is on, at intervals of 5 seconds. The tone level is controlled by a machine which selects at random one of seven levels available in steps of 2 db. Each level is presented five times. When the results of the test have been plotted to show at each level the number of times the tone was heard, and when a smooth curve has been drawn, one point on the curve indicates the level at which the tone is heard in 50 percent of the presentations, which level is accepted as the threshold level.

Another curve, closely related to the M.A.F. pressure level shown by curve 1 in Fig. 8, has been drawn to represent the minimum audible pressure level (M.A.P.) under an earphone receiver cap. In order to draw this second curve we have again started by raising the binaural zero loudness contour by 2.5 db to obtain the binaural M.A.F. of the articulation test crews. To this M.A.F. has been added the difference between the monaural M.A.P. and the binaural M.A.F. from Sivian and White.[12] The resulting M.A.P. for the articulation test crews is shown by curve 2 in Fig. 8. The corresponding technique for observing thresholds is a refined technique of the types described for curve 1.

To compare the measured acuity of the 1919–1920 articulation test crew with the acuity adopted as typical for all the crews, certain discrete points have been added to Fig. 8 for frequencies from 130 to 2000 c.p.s. These points were derived from measurements of the acuity of hearing of articulation crew members. In these tests a single frequency e.m.f. was applied by an oscillator connected in place of the microphone of the high quality system. The voltage was found which caused the tone to be at threshold for an observer listening in the customary manner to the sound from the receiver when the system gain was adjusted to give unity reproduction as described in the section of this paper dealing with responses. This voltage was expressed as an equivalent pressure applied to the microphone, the conversion being made by means of the thermophonic calibration of the microphone. The pressure so found for the different members of the crew were averaged at each frequency and the corresponding average pressure level was plotted as a discrete point in Fig. 8. The agreement between these points and the corresponding ordinates of curve 2 is regarded as sufficiently close to justify the adoption of the curves in Fig. 8 to represent the acuity of hearing of the 1919–1920 articulation crew.

The two curves in Fig. 8 together describe the first of the two desired measures of the acuity of the listeners in the articulation test crews—namely, the acuity as it would have been measured by use of a refined technique, the listeners being experienced observers. Another desired measure of the acuity of the same listeners is shown in Table VI. This table expresses the acuity of the typical listener in the articulation crews when the members of the crews had little or no previous experience in auditory testing. The acuity is given by two audiograms corresponding to the use, respectively, of the 2A audiometer and of the proposed A.S.A. standard audiometer. The origin of these audiograms, and the meaning of the other columns in Table VI, will now be explained.

For each of the twenty persons who acted as listeners in the 1928–1929 and 1935–1936 articulation tests, an audiogram is available which was measured by aid of the 2A audiometer using customary technique. In twelve instances the audiogram was taken either just before or shortly after the individual was employed by the Bell Telephone Laboratories. In the remaining eight instances, the person had been in the employment for from one-half to three years, but probably had not received training in auditory work to the degree represented by the later work in the articulation test crews.

Of the twenty listeners just mentioned seven took part in the loudness and masking tests during which threshold observations were made by the refined techniques previously described. In addition, for four other members of the loudness and masking crews there are audiograms taken either just before or shortly after the individual was employed by the Laboratories. So for twenty-four young adults who acted as listeners in the various crews and tests, audiograms are available which express the acuity of each listener when almost or quite inexperienced in auditory observations, each audiogram having been measured by the 2A audiometer with customary technique. The total of twenty-four persons thus considered together will be referred to as the group of 24.

At each frequency the average was found of the 24 values of hearing loss measured for these persons by means of the 2A audiometer. For each person the hearing loss of only one ear was used in finding these averages, namely the loss for the ear used customarily when listening with one earphone as in the articulation tests. The average hearing loss thus obtained at each frequency (the median, at 8192 c.p.s.) has been entered in column 2 of Table VI, and together these values constitute an audiogram which has been adopted as the 2A audiometer audiogram of the typical listener of the articulation test crews at a time when each listener had little or no experience in making auditory observations. This audiogram corresponds to the use of a dial attenuation step of 5 db and to the selection of that value of hearing loss which is indicated by the dial at the lowest setting at which the tone is definitely heard.

Typically the average value of hearing loss in column 2 of Table VI represents 24 observations distributed over a range of 15 to 25 db. Typically the root-mean-square deviation of one observation from the mean is about 5 db. The greatest value of the r.m.s. deviation is 6.9 db, for the observations at 8192 c.p.s. For this frequency column 2 shows the median instead of the mean hearing loss.

The values shown in column 3 of Table VI were derived from those in column 2 by aid of the differences between the tone pressure levels supplied by the two audiometers at zero dial setting. These values of hearing loss in column 3 together constitute the audiogram of the typical listener of the articulation test crew as measured by aid of the proposed A.S.A. standard audiometer using standard technique, the audiogram referring to a time when the listeners had little or no experience.

Column 4 of Table VI shows at each frequency the pressure level under the cap of the earphone of the 2A audiometer for the dial setting indicated by column 2. These pressure levels are derived from the search tube pressures measured by W. A. Munson and reported by Steinberg and Gardner.[18] To compare with these levels in column 4, column 5 shows at each of the 2A audiometer frequencies the M.A.P. level from curve 2 in Fig. 8. The difference between columns 4 and 5 of Table VI is shown in column 6 and will be discussed briefly.

From the manner of deriving columns 4 and 5 of Table VI it is seen that these two columns show two values of M.A.P. for the same listener and therefore one might expect them to be approximately equal instead of unequal by the rather large differences shown in column 6. However, these differences can be explained in the following manner. Column 4 shows the M.A.P. when the threshold adjustment is the lowest audiometer dial setting at which the tone is definitely heard. If the audiometer step had been much smaller than 5 db, these threshold levels would have been lower on the average by about one-half step or 2.5 db. The differences in column 6 would then have been smaller by 2.5 db (except at the lowest frequency). Moreover, the audiogram from which column 4 was derived applies to the typical listener when inexperienced, whereas column 5 applies to the listener when experienced. From a comparison of audiograms before and after obtaining experience in auditory observations, for eight members of the group of 24, it is concluded that a fair value for the average apparent decrease in hearing loss as measured by the audiometer is 4 or 5 db. This is in reasonably good agreement with a difference of about 6 db found by Steinberg and Munson[17] between the threshold levels at 1000 c.p.s. for a group of inexperienced observers and a group of experienced observers. Thus experience in auditory testing may be regarded as accounting for about 5 db of the difference shown in column 6 of Table VI.

After accounting as just indicated for about $2.5+5=7.5$ db of the total difference shown in column 6 at each frequency, there remains a residue of difference which may be attributed to the effect of using dissimilar techniques in the two sets of observations from which columns 4 and 5 were derived. Neglecting the two extreme frequencies, namely 128 and 8192 c.p.s., this residual difference ascribed to technique amounts to about 3 db on the average, which is regarded as a reasonable value. Thus it is concluded that the acuity of the inexperienced listener represented by the audiograms in column 2 and 3 of Table VI is consistent with the acuity of the same person as an experienced listener represented by the curves in Fig. 8.

In the last line of Table VI, in columns 2 and 3, the weighted average hearing loss for speech is entered as approximately or exactly equal to -4 db. The manner of obtaining this weighted average from the single frequency observations is explained in Appendix 2.

5. TALKING LEVELS AND PROFICIENCY FACTORS OF THE ARTICULATION CREWS

For a listener having a definite acuity of hearing and using a telephone system of chosen physical characteristics, the received speech has an intelligibility which depends upon the acoustic spectrum level produced at each frequency by the talker. Therefore we need to know these spectrum levels for the various crews involved in the basic articulation tests.

The talking level is an over-all measure of the acoustic level and is here defined to be the long-time average intensity level of speech at a point one meter directly in front of the talker's lips in a free acoustic field and is designated β_t. The long-time average intensity is the average taken over a length of time sufficient to include the typical pauses between syllables and words. The measurements of talking levels of the various articulation test crews involve the use of actual test syllables with the introductory phrases, or an approximate phonetic equivalent.

Measurement was not made of the spectrum level *versus* frequency curves for each of the talkers used in the articulation tests, but data are available for certain of these talkers and for other talkers who are considered typical of those used in the tests. These data, averaged, and smoothed,[3] are given in Table XXX of Appendix 1 in the column headed B_s. The talking level β_t corresponding to this spectrum level B_s is $\beta_t=68$ db from 10^{-16} watt/cm^2, which has been chosen as typical of conversational speech. Tests made upon talkers using the telephone indicate that for about 95 percent of them the talking level may be anywhere from 55 to 75 db. It will be seen from Table VII that the average talking level for any one of the various articulation test crews varied by not more than 2 db from this 68-db value. Therefore, it was considered that the spectrum level curves were raised or lowered uniformly by the difference between 68 and the observed talking levels. The values given in Table VII are those used in the calculations of articulation *versus* gain to be described.

If the shape of the spectrum level curve for a talker departs radically from that shown in Table XXX, for

[18] J. C. Steinberg and M. B. Gardner, "On the auditory significance of the term hearing loss," J. Acous. Soc. Am. **11**, 276 (1940), Fig. 9.

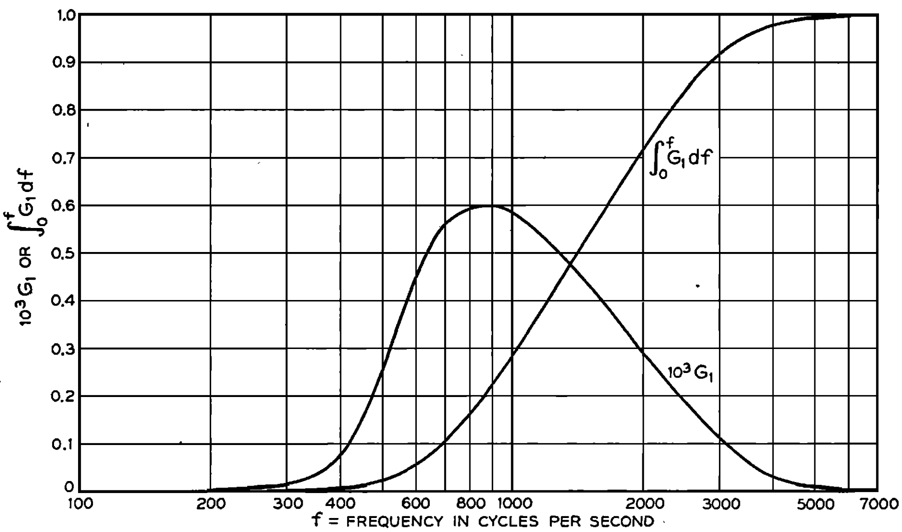


FIG. 9. The functions G_1 and $\int_0^f G_1 df$. G_1 is the frequency weighting factor for loudness, when speech is at the threshold of audibility. (See also Fig. 50 and Table VIII.)

example as in whispered speech, then in order to calculate articulation by the present method it is necessary to treat the case differently from any considered here. However, it is presumed that the departures in db at each frequency of such a spectrum level curve from the typical one may be treated as a corresponding change in the response curve of the transmission system.

Table VII includes also the values of the proficiency factor p for the various articulation crews. These values were obtained from the articulation tests in the following manner. As already stated, the proficiency factor $p=1.0$ was arbitrarily made to apply to the 1935–1936 testing crew in the tests of the high quality and filtered systems by choosing the value of Q in Eq. (3) equal to 0.55. The same proficiency factor was used for the other 1935–1936 systems.

In Fig. 4 a pair of curves was drawn to represent the maximum values of observed sound articulation s_{3M} for the low pass and high pass filters, respectively, in the 1935–1936 tests. In the same plot another pair of curves represents the corresponding quantity s_3 for the 1928–1929 filtered systems. At the intersection point of each pair of curves the sound articulation is the same and hence the values of articulation index for the low pass and the high pass filtered systems must be equal to each other so that $A=0.5$ for each of these systems. The intersection point is seen to be the same for the two pairs of curves. It follows that for a system having the index $A=0.5$ the 1928–1929 crew obtained the same value of sound articulation as did the 1935–1936 crew. For this reason the proficiency factor $p=1.0$ was applied also to the 1928–1929 crew and is so shown in Table VII.

Although the proficiency factor $p=1.0$ has here been applied to all of the 1928–1929 systems, there is some evidence that for two or three of the filtered systems the crew was relatively low in training.

For any particular testing crew it may happen that articulation data for noiseless filtered systems are not available to the extent required for a plot of the type of Fig. 4. In such a case the proficiency factor is chosen to fit the articulation values observed by the crew using one of the systems, preferably the system having highest observed articulations, and then this value of the factor is applied to all the other systems tested by the same crew. For example, this procedure has been followed in considering the fourth major group of articulation tests, which were made at Harvard University with due regard to the requirement that when conditions are changed, proficiency tests must be made until the training reaches the previous stable level. The value $p=1.0$, which has been used here for all of the tests of group IV as well as for those of groups II and III, represents apparently a typical degree of proficiency for a testing crew composed of selected normal young adults who are well trained and have an incentive to give attention to their work.

An examination of the results of the 1919–1920 articulation tests yields definite evidence of an increase in proficiency on the part of the testing crew during this long series of tests. This is shown by the observed values of articulation for the high quality system for values of gain near the optimum, as plotted in Fig. 14, where the maximum observed value of articulation was between 0.94 and 0.95 in November, 1919, but had risen to become 0.95 to 0.96 in January, 1920, and reached 0.975 in April, 1920. To make the calculated articulations for this high quality system agree approximately with observed articulations at optimum gain, the proficiency factor $p=0.88$ has been adopted for November, 1919, $p=0.98$ for January, 1920, and $p=1.0$ for February to December, 1920. Between these tests of the high quality system, many resonant systems and filtered systems were tested, so for these systems the proficiency factor has been arranged on the chronological basis derived from the tests upon the high quality system as just explained. Thus the resonant systems represented in Fig. 33 which were tested in early December, 1919 were assigned the value $p=0.9$; and for the resonant systems in Fig. 34 which were tested in late December, 1919 and early January, 1920, the assigned value was $p=0.94$.

The tests represented in Figs. 35 and 36 were made in 1924, and those in Fig. 37 were made in 1925. The group I high quality system was used in these tests, without and with various interfering single frequency tones admitted at the received levels indicated in the plots. For each of these figures the value of the practice factor has been assigned so as to make calculated articulations agree with those observed, for the high quality system with no noise and with optimum gain.

From the standpoint of talking level, it is convenient to discuss the four groups of articulation tests in the reverse of the chronological order.

During the Harvard tests, the talker endeavored to keep his talking level uniform through hearing his own voice and watching a VU-meter, monitoring undistorted speech, which had been calibrated by comparison with a square law integrator used under free field conditions. The talking level in these tests was reported by Egan and Wiener[4] as 70 db.

The articulation tests made at the Bell Telephone Laboratories in 1935–1936 involved the use of automatic equipment which has been described in published articles,[6] including two volume indicators monitoring, respectively, the undistorted speech output of the microphone and the input to the receivers necessarily following the distorting network, if any. The first of these instruments gave a visual signal. The second instrument gave a printed record of the average speech wave level actually applied to the receivers, thus including the small variations in average level which occurred in spite of the attempt of the talker to maintain a uniform over-all talking level through hearing his own voice and watching the visual signal.

When the response of a system is known from talker to volume indicator, the reading of the volume indicator may be used in calculating the acoustic talking level. From such calculations for the high quality system, the 1935–1936 talking levels in Table VII were derived, with an uncertainty of about ±1 db.

In the 1936–1937 articulation tests of systems having either a carbon microphone or the corresponding linear "hybrid" microphone, the acoustic speech wave was applied to the microphone by an artificial voice forming a part of the caller's control circuit.[6] The human talker spoke into the high quality microphone of the caller's control circuit. As the response of the entire system from human talker to recording volume indicator was known, the acoustic talking level could be calculated. The average talking level so calculated varied from about 66 to about 68 db for the different systems which were tested. Another measure of the talking level of this crew was made using a calibrated condenser microphone under approximately free field conditions, resulting in the value 66.5 db which has been used here for all of these systems.

During the Harvard articulation tests every talker attempted to speak at the same over-all level. Likewise in the 1935–1936 tests the same over-all level was attempted by every talker, and the records indicate that such uniformity was attained on the average to within a fraction of 1 db. Consequently the average talking level and also the average received speech level is a relatively precise quantity for any one of these tests. The technique was somewhat different in the 1928–1929 tests. By preliminary trials each talker established his or her natural voice level and thereafter tried to talk uniformly at that level, hearing his own voice and watching a volume indicator which monitored the undistorted speech. The average volume indicator reading for every talker was recorded in writing for each test, the spread of such readings with respect to the average for all eight talkers being frequently greater than ±6 db.

Although the volume indicator readings for the eight talkers differed considerably among themselves in the 1928–1929 articulation tests, the average of the eight readings for a test rarely differed by more than 1.5 db from the grand average for all the tests. While these differences may correspond to actual differences in the average acoustic talking level, they have here been used as shifts applied to the over-all gain setting so that the data could be plotted as though obtained with a uniform talking level. The grand average volume indicator reading was used together with the response from talker to volume indicator in calculating the average acoustic talking level for the entire group of tests. The response entering into this calculation included the average over-all efficiency of the various condenser microphones employed in the tests. Because this average efficiency was not so accurately known as in the later tests, it was thought desirable to supplement the calculation just mentioned by a second calculation based upon threshold observations of speech delivered

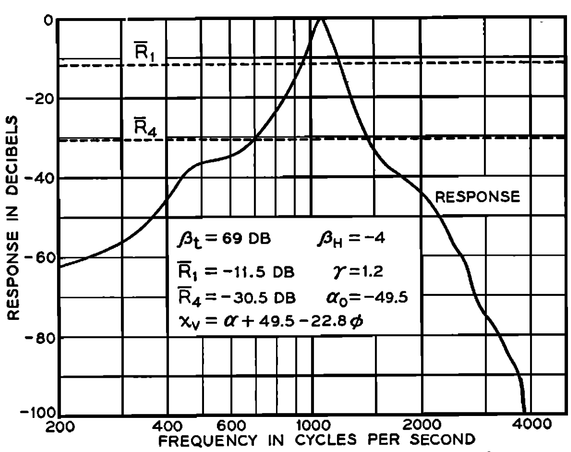


FIG. 10. Response of resonant system II–RN–1060, with weighted averages $\bar{R}_1$ and $\bar{R}_4$ and values of γ, α_0 and x_V.

by a system whose response characteristics are known. As will be seen later, Eq. (22) gives the value of the talking level in terms of a weighted average of the response $\bar{R}_1$ and the gain α_0 necessary to deliver the speech so it will be at the threshold of hearing for the reference crew listeners. This second calculation used the 1928–1929 observations of the gain adjustment which caused speech to be at the threshold of audibility when heard over the high quality system. The two calculations were in satisfactory agreement and together resulted in the talking level 69 db in Table VII, with an uncertainty probably less than ±2 db.

In the 1919–1920 articulation tests, no direct measurements of talking level were made with volume indicators or otherwise, nor were there any direct observations of the gain adjustment for the threshold level of received speech. However, the threshold adjustment was inferred from the plot of syllable articulation *versus* attenuator setting for the high quality system, through an extrapolation required to obtain the intercept of the curve drawn through the observations. The gain setting of system I (1919–1920 high quality) thus obtained for speech at threshold corresponds to the abscissa $\alpha_0 = -61.0$ db in Fig. 14, with an uncertainty arising from the extrapolation of probably less than ±3 db. Using this value of α_0 together with Eq. (22) just mentioned, the talking level for system I was found to be 69 db, which is entered in Table VII as an estimated rather than a measured value.

6. FORMULATION OF THE GENERAL METHOD OF CALCULATING THE ARTICULATION INDEX *A*

The articulation index depends upon (1) the response characteristic *R versus f* for the system; (2) the over-all gain α from this response curve; (3) the noise conditions at the listener's ear; and (4) other special types of distortion such as overloading, room reverberation, etc.

TABLE VIII. Values of f *vs.* $\int_0^f G_1 df$. (See also Figs. 9 and 50.)

$\int_0^f G_1 df$	0.005	0.015	0.025	0.035	0.045	0.055	0.065	0.075	0.085	0.095
0.00	400	470	*510*	540	570	600	620	*640*	660	680
0.10	700	715	*730*	750	770	785	800	*820*	835	850
0.20	870	885	*900*	920	935	950	965	*980*	1000	1015
0.30	1035	1050	*1070*	1085	1105	1120	1140	*1160*	1175	1195
0.40	1210	1230	*1250*	1270	1290	1320	1340	*1360*	1380	1400
0.50	1420	1450	*1470*	1490	1520	1540	1560	*1590*	1610	1630
0.60	1660	1690	*1710*	1740	1770	1800	1830	*1860*	1890	1930
0.70	1960	1990	*2020*	2060	2090	2130	2170	*2210*	2250	2300
0.80	2340	2380	*2430*	2480	2530	2590	2640	*2700*	2760	2830
0.90	2910	3000	*3100*	3200	3330	3490	3660	*3930*	4420	5400

TABLE IX. Values of f *vs.* $\int^f G_4 df$.

$\int_0^f G_4 df$	0.005	0.015	0.025	0.035	0.045	0.055	0.065	0.075	0.085	0.095
0.00	180	250	*320*	370	410	440	470	*500*	530	555
0.10	580	605	*630*	655	680	705	730	*750*	775	800
0.20	825	845	*870*	895	920	945	965	*990*	1015	1040
0.30	1065	1095	*1120*	1145	1175	1200	1230	*1260*	1290	1320
0.40	1350	1380	*1410*	1440	1470	1505	1535	*1570*	1605	1640
0.50	1680	1720	*1760*	1800	1840	1880	1920	*1960*	2000	2040
0.60	2090	2140	*2190*	2240	2290	2340	2390	*2440*	2500	2550
0.70	2610	2670	*2740*	2810	2880	2950	3020	*3090*	3160	3250
0.80	3340	3430	*3520*	3620	3720	3830	3950	*4080*	4210	4350
0.90	4510	4690	*4890*	5100	5350	5660	6060	*6600*	7430	9600

TABLE X. Articulation *vs.* gain for ideal system.

(1) $\alpha - \alpha_0$ or x_V or x_E	(2) S_{23}	(3) S_3	(4) s_{3M}	(5) A	(6) V	(7) E
0	0	0	0	0	0	1.0
5	0.073	0.024	0.189	0.05	0.05	1.0
10	0.163	0.084	0.367	0.11	0.11	1.0
15	0.289	0.216	0.567	0.20	0.20	1.0
20	0.437	0.382	0.715	0.30	0.30	1.0
25	0.604	0.563	0.821	0.41	0.41	1.0
30	0.752	0.725	0.896	0.54	0.54	1.0
35	0.850	0.839	0.937	0.66	0.66	1.0
40	0.905	0.900	0.957	0.75	0.75	1.0
45	0.943	0.939	0.969	0.83	0.83	1.0
50	0.961	0.959	0.976	0.89	0.89	1.0
55	0.975	0.967	0.980	0.94	0.94	1.0
60	0.975	0.973	0.983	0.98	0.98	1.0
65	0.978	0.976	0.985	1.0	1.0	1.0
70	0.978	0.976	0.985	1.0	1.0	1.0
75	0.976	0.974	0.984	0.99	1.0	0.99
80	0.975	0.973	0.983	0.98	1.0	0.98
85	0.972	0.970	0.982	0.96	1.0	0.96
90	0.967	0.965	0.980	0.93	1.0	0.93
95	0.963	0.961	0.977	0.90	1.0	0.90
100	0.956	0.953	0.974	0.87	1.0	0.87
105	0.950	0.946	0.972	0.85	1.0	0.85
110	0.943	0.939	0.969	0.83	1.0	0.83
115	0.935	0.930	0.966	0.81	1.0	0.81
120	0.926	0.921	0.963	0.79	1.0	0.79

It has been found that A can be calculated as the product of four factors: the articulation growth factor V, the ear desensitizing factor E, the maximum articulation index factor F, and the special distortion factor H, or

$$A = V \cdot E \cdot F \cdot H. \quad (19)$$

A formula such as Eq. (19), which is empirical, can be made to fit the complex set of experimental data because one has considerable latitude in choosing the variables for determining each factor. The separation into factors of this type is useful for engineering purposes because it makes clear the effect upon A of changing the two important variables, namely, the shape of the response curve and the amplification or attenuation in the system.

The first factor V shows how A grows as the gain α in the system increases and it depends upon the effective gain x_V, which will be defined later.

The second factor E is dependent upon the level of the speech above the threshold level in the absence of noise—that is, the stimulation level in the listener's ear. It is unity when this level is below 68 db and gradually decreases as the level increases above this value.

The third factor F is dependent only upon the shape of the response curve. It is unity when the response curve is flat, and is between zero and unity for any other shape.

The fourth factor H is unity except when special types of distortion are present such as overloading, carbon microphone distortion, and when high intensity noise is present.

7. THE EFFECTIVE GAIN x_V

For any telephone system, the factor V in Eq. (19) grows from zero to the maximum value unity as the received speech level rises from the threshold level for audibility to some higher level. When the telephone system is an ideal flat response system, or an ideal filter system, the gain must be increased by 68 db in order that V may rise from zero to unity, and for such an ideal system the factor V may be expressed as a function solely of the level of the received speech above threshold. For many other types of system, however, V cannot be so simply expressed. In general, it has been found necessary to define a quantity called the *effective gain* x_v, which may be regarded as the *effective* level of the received speech above threshold, so that when x_V is known the value of V is determined. The effective gain is defined in terms of the following quantities: the actual gain α of the system, the response characteristic R *versus* f from which α is measured, the talking level β_t, and the hearing loss β_H of the listener. The following equation defines this effective gain.

$$x_V=\alpha+\beta_t-\beta_H-12+\bar{R}_1-\phi\gamma(\bar{R}_1-\bar{R}_4). \quad (20)$$

The first term α is the actual gain in db in the system from the responses shown on the plot of R *versus* f. The talking level β_t has been defined. The value β_H is the listener's hearing loss for speech. It was found that the average intensity level of undistorted speech was β_H+12 db when at the threshold for a listener having a hearing loss β_H. For the typical listeners in our tests $\beta_H=-4$ so that for all the articulation test crews $\beta_H+12=8$. Therefore, the level of the speech above threshold as it was uttered is $\beta_t-\beta_H-12$.

The values $\bar{R}_1$ and $\bar{R}_4$ are weighted average values of the response R. The method of deriving them will be explained later.

The coefficient ϕ depends directly upon x_V and becomes zero when x_V is zero and is equal to unity when x_V is 40 db or greater.

The coefficient γ depends upon the shape of the response curve and lies between zero and 1.4.

It will be seen that for an ideal system where $\bar{R}_1=\bar{R}_4=R$, the value of x_V becomes

$$x_V=(\beta_t-\beta_H-12)+(\alpha+R). \quad (21)$$

The first term is the level above threshold level of the speech as uttered and the second term is the amplification given to the speech by the system. In other words, the value of x_V for an ideal flat response system is the db above threshold level of the speech at the listener's ear. When x_V is equal to zero, then ϕ in Eq. (20) is also equal to zero. Let α_0 be the amplification under these conditions; it is the db gain necessary to bring the speech delivered by the system to the threshold level of the listeners. Then, by Eq. (20),

$$\alpha_0=\beta_H+12-\beta_t-\bar{R}_1. \quad (22)$$

The value 12 was obtained from threshold measurements of speech using an approximately flat response system. Experimental values of α_0 were obtained for systems II and III (response curves shown in Figs. 15 and 16) for talkers having $\beta_t=69$ db and listeners having $\beta_H=-4$ db, and values of $\bar{R}_1$ calculated from the response curves. It was found that these values satisfied Eq. (22) when the constant was 12 db.

It should be emphasized that for the crews used in the articulation tests reported in this paper $\beta_H=-4$ db so for these crews

$$\alpha_0=8-\beta_t-\bar{R}_1.$$

For the tests made at Harvard the value of β_H could depart considerably from -4 db without affecting the calculated results since the threshold levels were determined by the noise. Equation (22) is very important because it enables one to calculate the gain α_0 required to reach the threshold level of hearing, that is the gain at which speech is just detectable as determined by the technique described in Appendix 1. This gain α_0 is regarded in the present paper as the foot of the articulation *versus* gain curve, neglecting any articulation scoring which in an actual test may occur through correct guesses even when no speech sounds are heard.

In order to use Eq. (22) one must know how to calculate $\bar{R}_1$. Loudness studies have shown that near threshold levels $\bar{R}_1$ can be obtained by the equation

$$10^{\bar{R}_1/10}=\int_0^\infty G_1 10^{R/10}df. \quad (23)$$

The value $10^{R/10}df$ is proportional to the speech power carried by the frequency band df. Thus $10^{\bar{R}_1/10}$ is a weighted average value of the speech power. The value of the weighting factor G_1 depends both upon the hearing and upon the speech characteristics. But it can be determined directly from threshold measurements of speech as follows.

Let α_0 be the amplification in an ideal system which delivers speech to the ear at the threshold level. If an ideal low pass filter is introduced having a cut-off frequency f_c, then the amplification must be increased $\Delta\alpha$ above α_0 for the speech to be again at the threshold level. Then, as shown in Appendix 1,

$$\int_0^{f_c} G_1 df=10^{-\Delta\alpha/10}. \quad (24)$$

Similarly for high pass filter systems

$$\int_{f_c}^{\infty} G_1 df=10^{-\Delta\alpha/10}. \quad (25)$$

From experimentally determined values of $\Delta\alpha$ the curve for $\int_0^f G_1 df$ in Fig. 9 was determined as outlined in Appendix 1. The values of G_1 were obtained from the slope of this curve when plotted with both coordinate scales linear as in Fig. 50.

Consider system II–RN–1060 which has a non-uniform response shown in Fig. 10, and let $\bar{R}_1$ be the calculated value as given by Eq. (23) and indicated in the figure. Also consider an ideal system having a uniform response equal to $\bar{R}_1$. Then the attenuation in db to bring the received speech to the threshold level will be the same in both the ideal and the peaked systems. At levels near the threshold the speech delivered in each system will sound equally loud.

However, when gains of more than 40 db above the threshold level are introduced in each system, the speech delivered in each case will no longer sound equally loud; the speech from the ideal flat response system will be the louder. Now it might be expected that when the loudness of the speech in both cases is the same, the effective gain would be approximately the same. If this is true, then at these levels $\phi=1.0$ and γ is approximately equal to unity (see Eq. (20)) so that

$$x_V=\bar{R}_4+\alpha+(\beta_t-\beta_H-12)$$

and system II–RN–1060 must have a weighted average response equal to $\bar{R}_4$ instead of $\bar{R}_1$ to make the two effective gains equal. However, we know from our loudness measurements upon speech (see Appendix 1) that at these higher levels one must take a weighted average of the fourth root of the speech power to obtain equality of loudness, or

$$10^{\bar{R}_4/40}=\int_0^\infty G_4 10^{R/40} df. \tag{26}$$

It is shown in Appendix 1 that G_4 is related to G_1 through the properties of speech and hearing, so that

$$G_4=G_1^{\frac{1}{4}}10^{-3\kappa/40} \quad \text{(constant)} \tag{27}$$

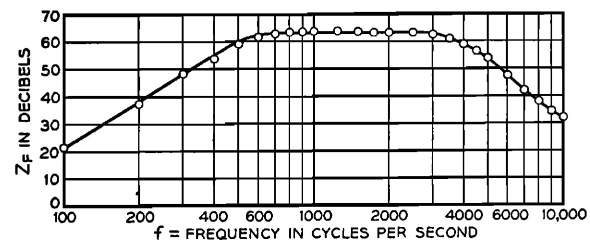


FIG. 11. Level Z_F of each critical band of undistorted speech in decibels above threshold, when the speech is at the optimum level for interpretation.

TABLE XI. $W(x_W)$ *vs.* x_W.

x	0	1	2	3	4	5	6	7	8	9	Difference
0	1.0	0.997	0.994	0.990	0.985	0.980	0.973	0.966	0.958	0.950	0.007
10	0.940	0.930	0.920	0.910	0.899	0.887	0.874	0.860	0.846	0.832	0.012
20	0.818	0.804	0.789	0.774	0.759	0.744	0.728	0.712	0.695	0.678	0.017
30	0.660	0.642	0.623	0.603	0.582	0.561	0.539	0.516	0.492	0.467	0.022
40	0.441	0.415	0.390	0.365	0.340	0.315	0.291	0.267	0.244	0.222	0.022
50	0.202	0.183	0.165	0.148	0.132	0.118	0.104	0.091	0.080	0.070	0.013
60	0.060	0.050	0.040	0.030	0.022	0.015	0.010	0.005	0.000		0.010

where κ is the critical band width in db and the constant is determined by the condition that

$$\int_0^\infty G_4 df=1.0. \tag{28}$$

Values of G_4 thus determined are given in Table XXX of Appendix 1. The values of $\int_0^f G_1 df$ and $\int_0^f G_4 df$ are given in Tables VIII and IX. The tables are arranged so that the frequency f is given for each 0.01 increase in the integral since this form is most useful for calculation. To evaluate Eq. (23) one plots $\int_0^f G_1 df$ as abscissas and $10^{R/10}$ as ordinates, then the area under such a plot is the required value of the integral. It is seen that an average value of $10^{R/10}$ taken at the one hundred different frequencies given in Table VIII gives a very good approximation to this value. For most systems it is sufficiently accurate to use only the twenty frequencies in italics. The other frequencies are used only where R is changing rapidly with frequency. The calculation of $\bar{R}_4$ is made in a similar manner using Table IX. The value of $\bar{R}_4$ for system II–RN–1060 is shown in Fig. 10. A flat response system having this response $\bar{R}_4$ will deliver speech which at the higher levels sounds equally loud to that received from system II–RN–1060.

Equation (20) for the effective gain x_V can be written

$$x_V=\alpha-\alpha_0-\phi\gamma(\bar{R}_1-\bar{R}_4) \tag{29}$$

the value of α_0 being that in Eq. (22). This shows that for systems where γ is equal to zero (ideal filter systems) and for systems where $\bar{R}_1=\bar{R}_4$ (ideal flat response systems) the effective gain is equal to $\alpha-\alpha_0$ or the db above threshold level. For other systems the effective gain is always less than $\alpha-\alpha_0$ by the amount $\phi\gamma(\bar{R}_1-\bar{R}_4)$. For some systems the articulation growth curve approximately follows the speech loudness growth curve for a considerable range of levels above the threshold level of received speech, but this statement is not true for telephone systems in general.

It remains, then, to determine ϕ and γ before the value of the effective gain x_V can be calculated for any system. Before doing this it is necessary to determine V, E and F because these functions are involved in the determination of ϕ and γ.

8. DETERMINATION OF V AND E

The articulation growth factor V and the ear desensitizing factor E are determined in the following manner from articulation data on three different systems designated I–III, for which the responses are given in Figs. 14–16.

It is seen that these systems had only an approximately flat response so certain small corrections, to be discussed later, are applied to obtain results corresponding to an ideal flat response system. The final results are given in Table X.

For the ideal system, $\bar{R}_1=\bar{R}_4=0$ so that by Eq. (29)

$x_V = \alpha - \alpha_0$. Thus in Table X the level above threshold $\alpha - \alpha_0$ and the effective gain x_V are both given by column 1. So also is the argument x_E of the function E, as will be explained.

In columns 2–4 are given the articulation values S_{23}, S_3 and s_{3M} corresponding to the values of $\alpha - \alpha_0$ in column 1. For each of these articulations the value of the articulation index A has been found by the relations in Table III and Fig. 7. The average of the three such values of articulation index, for each value of $\alpha - \alpha_0$, is given in column 5.

For the ideal system the factors F and H are each equal to unity. So for Table X, by Eq. (19)

$$A = VE.$$

Thus for each value of A in column 5 we must find a value of V and of E such that the product VE is equal to A. By an arbitrary choice, the factor E has been taken as unity for all values of $\alpha - \alpha_0$ from 0 to 68 db, and the factor V has been taken as unity for $\alpha - \alpha_0 \geq 68$ db. Thus the values of V and E in columns 6 and 7 were obtained.

It is considered that V becomes less for values of $\alpha - \alpha_0$ below 68 db because more and more of the components of speech drop below the threshold of hearing as the level is lowered until all the speech components are below the threshold when $\alpha - \alpha_0$ becomes equal to zero. Therefore, the value of V is determined in terms of x_V by Table X or

$$V = V(x_V).$$

Above $\alpha - \alpha_0 = 68$ db another cause is operating to reduce the articulation as the gain is increased, namely, the sounds become so loud that the ear is fatigued by the loud sounds and cannot differentiate as accurately the succeeding softer sounds. Consequently, E is considered to be dependent upon the db above the threshold level when no noise is present. This corresponds to the stimulation level of the nerves of hearing. When noise is present this level is slightly increased but is raised only 3 db when the db above threshold for the noise and the speech are equal. For these reasons it is considered sufficiently accurate to regard E as dependent upon $\alpha - \alpha_0$ only, where α_0 in this case is the gain adjustment for threshold for the condition of no noise in the system. Thus E is the same for systems in the quiet or for any amount of noise. Therefore,

$$x_E = \alpha - \alpha_0$$

and

$$E = E(x_E) \qquad (30)$$

so for this ideal system $x_V = x_E = \alpha - \alpha_0$ but this is not true for other systems. The function E is considered unity for values of x_E from 0 to 68 db but determined by the values in Table X above 68 db. The corresponding values of x_E and E are given in columns 1 and 7 of Table X.

It should be realized that the decrease in the factor E, as the received speech level is increased above the level $\alpha - \alpha_0 = 68$ db, has been determined largely by the earliest group of articulation tests. The later tests cited here, upon relatively flat response systems, did not reach sufficiently high levels to confirm or deny these results. Some more recent articulation tests have been interpreted as indicating that for some observers there is little or no decrease in articulation at high received speech levels. However, the use of the factor E as given in Table X, with a droop at high levels, has resulted in a better over-all fit of calculated articulations upon observed articulations for all the systems tested than would have been obtained by the present method without such a function.

9. MAXIMUM ARTICULATION FACTOR *F*

The effect of a change of amplification upon the calculated articulation index of a telephone system is accomplished through the factors V and E, and sometimes through the factor H, in Eq. (19). The factor F is not dependent upon the gain of the system, hence this factor sets a limit which the articulation index cannot exceed but which it can equal if each of the other factors has the value unity.

When the gain α of the system is adjusted so that the effective gain x_V is equal to 68 db, then as has already been explained the factor V is equal to unity. Let this particular gain be designated as α_F. This is the gain that gives the condition for calculating F. To express the value of α_F, it is evident that at such levels ϕ is equal to unity so that from Eq. (29)

$$\alpha_F = \alpha_0 + 68 + \gamma(\bar{R}_1 - \bar{R}_4). \qquad (31)$$

The factor F depends upon the relative response at the various frequency regions. Its value is given by the equation

$$F = \int_0^\infty D \cdot W \cdot df \qquad (32)$$

where D is such a function of frequency that the product Ddf is equal to the element dA of articulation index carried by the frequency region between f and $f+df$ when this region is at the optimum level for speech interpretation. This function D is the frequency importance function for articulation which was defined by Eq. (13) and has the values given in Fig. 5.

The factor W which multiplies the function D in Eq. (32) is a quantity having any value from zero to unity. This factor W determines the reduction of dA due to the interval df being sent to the ear at levels below the optimum level for speech interpretation. Ideally W would be determined from articulation tests of a system which could be so altered that the received level of any chosen frequency interval could be set at any desired value, with no change in other frequency regions except an over-all adjustment for the optimum.

The factor W should not be confused with the factor V which ideally is determined from articulation tests

TABLE XII. Values of f vs. $\int_0^f Ddf$ from curve in Fig. 5.

$\int_0^f Ddf$	0.005	0.015	0.025	0.035	0.045	0.055	0.065	0.075	0.085	0.09
0.00	200	260	*310*	350	385	415	470	*470*	500	530
0.10	555	585	*610*	635	660	685	715	*740*	770	800
0.20	825	855	*880*	910	935	965	990	*1020*	1050	1080
0.30	1110	1140	*1170*	1200	1230	1270	1300	*1330*	1370	1410
0.40	1440	1480	*1520*	1550	1590	1640	1680	*1720*	1760	1810
0.50	1850	1900	*1950*	1990	2040	2090	2140	*2200*	2250	2300
0.60	2360	2410	*2470*	2530	2580	2640	2700	*2770*	2830	2890
0.70	2960	3020	*3090*	3160	3230	3310	3390	*3480*	3560	3640
0.80	3730	3820	*3920*	4020	4120	4230	4350	*4480*	4610	4740
0.90	4890	5050	*5220*	5400	5610	5800	6060	*6370*	6750	7300

of a flat response system having only such gain adjustments that all frequencies are raised or lowered equally.

Although the phenomena which control the factor W are different from those which control the factor V, the same range of magnitudes (namely 68 db) has been assigned to the two arguments upon which these two factors respectively depend, as the range which corresponds to the change of the factors from zero to unity. Thus W is a function of x_W such that $W=1.0$ when $x_W=0$ db, and $W=0$ when $x_W=68$ db. However, the growth of W from zero to unity as x_W changes is not the same as that of V as x_V changes.

The manner in which x_W is defined, and the relationship between W and x_W will now be described.

Let $\bar{R}$ be defined thus:

$$\bar{R}=\tfrac{1}{2}(\bar{R}_1+\bar{R}_4), \quad (33)$$

where $\bar{R}_1$ and $\bar{R}_4$ are as before the weighted average values of the response, using weightings which are appropriate respectively for very low received speech levels and for levels 50 db or more above threshold. Thus $\bar{R}$ is a sort of weighted average response for the whole range of levels.

It has been found satisfactory to assume that W

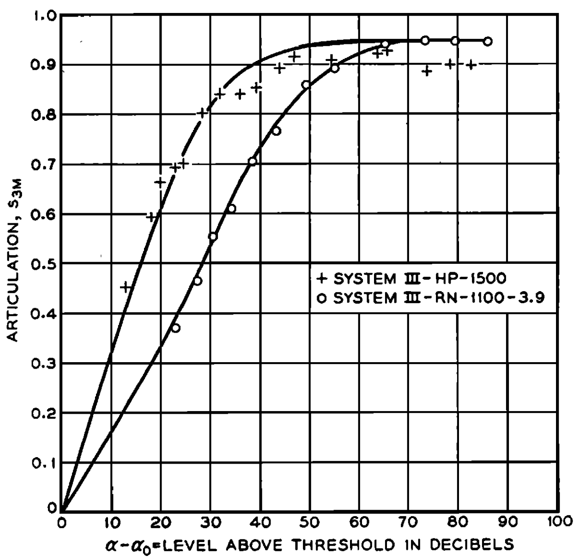


FIG. 12. Articulation vs. level of received speech above threshold, for a resonant system and a filter system having approximately equal maxima of articulation.

has the value unity whenever $R \geq \bar{R}$, and that W becomes less than unity as R becomes less than $\bar{R}$. When R decreases to a critical value R_c, W becomes zero and remains zero when the response is decreased further.

The variable x_W which determines W is defined by the equation

$$x_W=68(\bar{R}-R)/(\bar{R}-R_c). \quad (34)$$

The form of this equation was chosen so that when $R=\bar{R}$, $x_W=0$ db and when $R=R_c$, $x_W=68$ db. The difference $\bar{R}-R_c$ turns out to be equal to 68 db for the greater part of the frequency range. For this reason the constant 68 has been introduced into Eq. (34) so that for the greater part of the frequency range x_W becomes equal to $\bar{R}-R$, simplifying the calculation.

The value of W is taken equal to unity for $R \geq \bar{R}$, that is when x_W is zero or negative; and W is taken equal to zero for $x_W \geq 68$ db. It remains to find the form of W as a function of x_W, and also to find R_c as a function of frequency.

First, consider R_c. Measurements upon typical undistorted speech which is at the optimum level for interpretation, namely 68 db above the threshold level, show that the level Z_F of each critical band (in db above threshold) is that given in Fig. 11. The values were calculated from Eq. (41) given below. It is seen that from 700 to 3000 c.p.s. this level for each critical band is 63 db, which is 5 db less than that for speech as a whole. Thus it follows that statistically about three or four bands are cooperating together at one time to increase the effective level at threshold about 5 db over that of each band acting separately. It is thus seen that each band must be at least 5 db below its threshold level before it ceases to contribute toward the articulation index. Therefore, the value of R_c is given by

$$\bar{R}-R_c=Z_F+5. \quad (35)$$

Consequently, the values of $\bar{R}-R_c$ are 5 db greater than the ordinates of Fig. 11. So it is seen that Eq. (34) reduces to

$$x_W=68(\bar{R}-R)/(Z_F+5) \quad (36)$$

The values of W which correspond to values of x_W between 0 and 68 were determined empirically from the articulation data and are given in Table XI.

10. THE EFFECT UPON THE F FACTOR DUE TO THE MASKING OF ONE SPEECH SOUND BY ANOTHER

The masking effects of one speech sound by another are of two kinds. The first kind is due to the fatigue effect upon hearing lasting after the stimulus is gone. Its principal effect is in the same band of frequencies as that for the stimulating speech sounds, but it also has an effect, although much smaller, upon adjacent bands. This effect is presumably taken care of by the V and E factors. The second kind of effect is due to the masking action of one component of a speech sound upon a second component in a different frequency band.

In this case both components are sounding simultaneously. It is estimated that about one-third of the fundamental speech sounds are essentially in one frequency region and consequently need no correction for simultaneous masking. It is assumed that these are the unvoiced sounds and the long vowels ū (too), ō (toe), ȯ (not) and ȧ (father). The other two-thirds of the total speech sounds are assumed to have two or more important components. For this second group we will now calculate this second kind of masking. If B is the spectrum level of the noise, κ the critical band width in db, and β_0 the threshold level for pure tones, then the experiments on masking by thermal noise show that the masking M is equal to the level Z in db above threshold for the critical bands and can be computed correctly by the formula

$$M=Z=B+\kappa-\beta_0 \qquad (37)$$

except for the following conditions: (1) Except when dZ/df (absolute value) exceeds critical values which depend upon the frequency range and masking level; (2) except for very narrow frequency bands; (3) except for sharply resonant peaks for B; and (4) except for values near the threshold. The quantity Z is always taken equal to the value on the right-hand side of this equation. But M is equal to Z only with the exceptions noted.

These exceptions seem complicated but the following simplifications can be made without too much sacrifice in accuracy. The last restriction for values near the threshold can be removed because, for the condition for calculating F, these low levels never occur for frequencies that cause masking. Due to the statistical nature of speech it will be only a fraction of the time that speech energy will lie in a very narrow band or in a sharply resonant peak. Since masking occurs principally when voiced sounds are used, we can consider the components for the voices of men and women as spaced 180 cycles apart as an average. Consequently, in a frequency region where the Z *versus* $\log f$ curve has a peak, the curve is regarded as flat over a band 180 cycles wide and the ordinate is taken to be the average ordinate over the 180-cycle band. The same also applies to filters having band widths less than 180 cycles, although their case is academic, since there are no articulation data on such narrow band systems. Let a curve of Z *versus* $\log f$ be plotted and the peaks corrected as above. Then whenever the positive slope of the curve thus plotted is greater than 80 db per octave, the masking curve is higher by an amount ΔZ than the Z curve calculated by Eq. (37). One obtains a good

TABLE XIII. Values of $\Delta\alpha$ for determining ϕ.

$x_V=\alpha-\alpha_0=$	0	5	10	15	20	25	30	35	40	50
$\Delta\alpha=$	0	5	10	12.5	14.2	16	17	17	17	17
$\phi_\gamma=$	0	0.37	0.74	0.92	1.04	1.17	1.25	1.25	1.25	1.25
$\phi=\phi_\gamma\div 1.25=$	0	0.29	0.59	0.74	0.83	0.94	1.0	1.0	1.0	1.0
Adopted values of $\phi=$	0	0.22	0.45	0.65	0.85	0.93	0.97	0.99	1.0	1.0

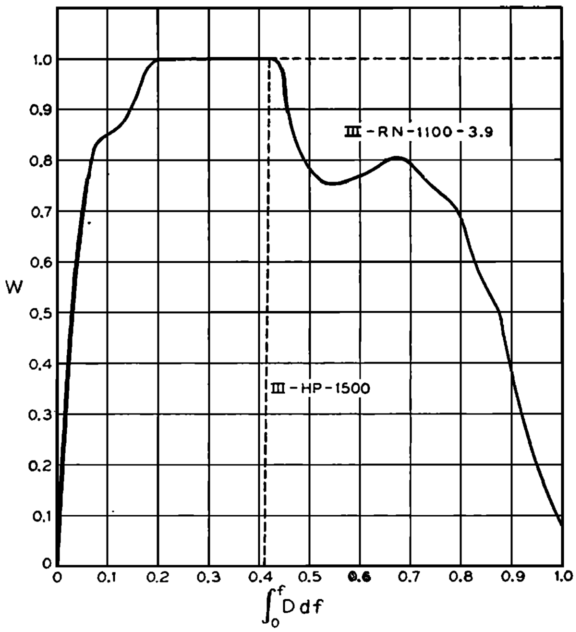


FIG. 13. The factor W *vs.* $\int_0^f Ddf$ for the two systems in Fig. 12.

approximation for the masking curve sloping toward the low frequency side of the curve by proceeding as follows. Let the point where the Z curve begins to have a slope greater than 80 db per octave be (f,y) where f is the frequency and y the ordinate on the masking curve. Draw a straight line from (f,y) to $(0.7f,y-40)$. Then from this last point to $(0.5f,y-60)$ and then to $(0.35f,y-70)$, and so on, halving the slope each half-octave as we go to the lower frequencies. The slope of the first half-octave is always 40 db per half-octave and does not change with different values of y. This geometrical construction also fits approximately the masking curve produced by a pure tone. This constructed series of straight lines will be referred to as the speech masking curve on the low frequency side.

On the high frequency side of a point on the Z *versus* $\log f$ curve the critical slope σ depends upon the ordinate y of this curve; but the curve M *versus* $\log f$ may be represented approximately by a single tangent line. This tangent line is drawn at the point where the slope of the Z *versus* $\log f$ curve exceeds the slope σ and the tangent continues with this critical slope σ. The masking data indicate that the following simple relation holds approximately, namely,

$$\sigma=75-Z_T/2 \qquad (38)$$

TABLE XIV. Values of ψ *vs.* x_γ.

x_γ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	1.20	1.20	1.20	1.20	1.20	1.20	1.198	1.196	1.194	1.192
4	1.185	1.176	1.167	1.158	1.149	1.140	1.128	1.114	1.097	1.083
5	1.070	1.060	1.045	1.030	1.010	0.990	0.970	0.940	0.905	0.870
6	0.840	0.805	0.770	0.740	0.710	0.685	0.655	0.625	0.595	0.560
7	0.530	0.500	0.475	0.445	0.415	0.390	0.360	0.335	0.310	0.285
8	0.265	0.240	0.215	0.185	0.155	0.130	0.110	0.090	0.070	0.055
9	0.045	0.035	0.025	0.015	0.008	0.003	0	0	0	0

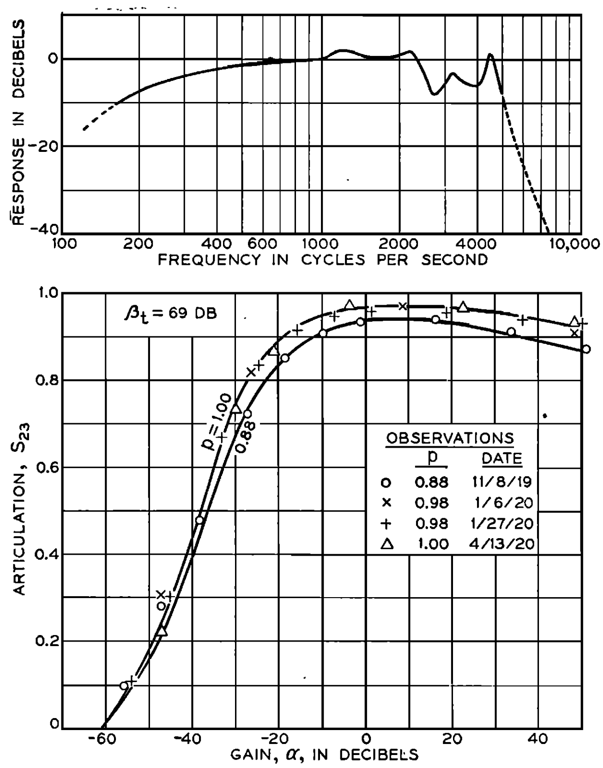


Fig 14

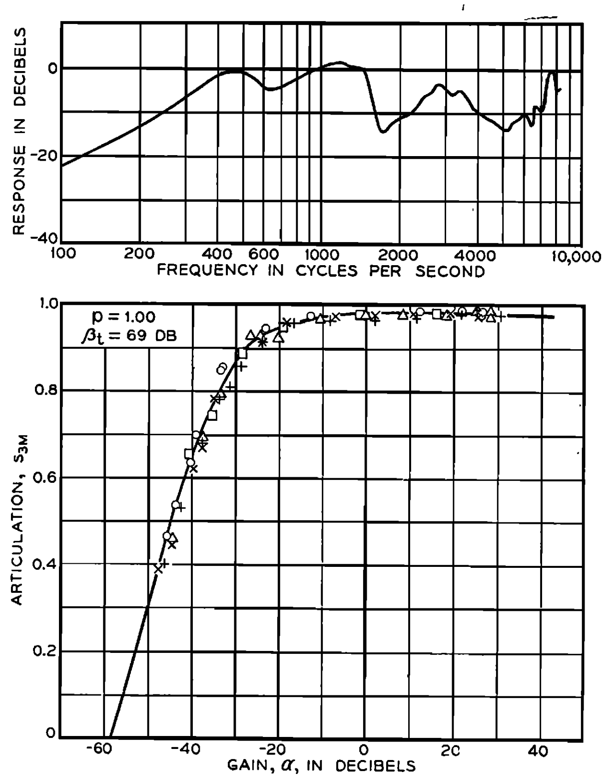


Fig 16

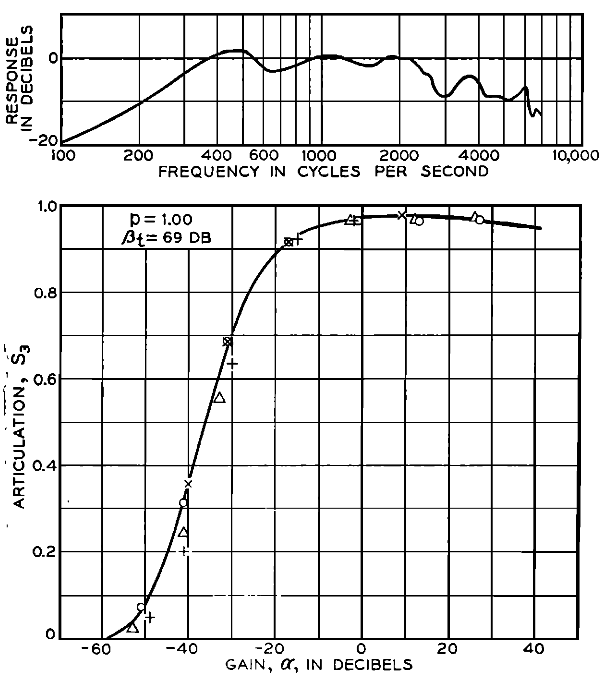


Fig. 15

FIG. 14. System I: 1919–1920 high quality system.
FIG. 15. System II: 1928–1929 high quality system.
FIG. 16. System III: 1935–1936 high quality system.
NOTE.—In Figs. 14–16 and 21–46 the response *vs.* frequency characteristics and the noise, if any, are shown in the upper plots and the articulation *vs.* gain observations (points) and calculations (curves) in the lower plots.

where Z_T is the value of Z computed from Eq. (37) at the point where the tangent is drawn, and σ is expressed in db per octave.

If these relations are applied to speech, then a curve will be derived which will be the db above threshold for each critical band of speech. This will also be the masking curve except where the slope exceeds the critical values discussed above. For speech transmitted through a system with response R and gain α from talkers having a talking level β_t, the spectrum level at

TABLE XV. Values of f and Δ.

$f=$	310	470	610	740	880	1020	1170	1330	1520	1720
$\Delta=$	−0.03	−0.01	0.01	0.03	0.04	0.04	0.03	0.02	0.00	−0.02
$f=$	1950	2200	2470	2770	3090	3480	3920	4480	5520	6370
$\Delta=$	−0.04	−0.05	−0.06	−0.06	−0.05	−0.04	−0.02	0.00	0.01	0.03

the listener's ear is given by

$$B=B_s+\Delta B_s+R+\alpha+(\beta_t-68) \tag{39}$$

where B_s is the average spectrum level for speech at one meter's distance from the lips of the speaker having a talking level of 68 db. The quantity ΔB_s is the peak level above B_s. (See Appendix 1.)

When the typical articulation crew member was listening to speech, the threshold level β_0' for each critical band was somewhat higher than the pure tone threshold level β_0, in the region of lower frequencies, so that $\beta_0'=\beta_0+\Delta\beta_0$. Here β_0' is the r.m.s. pressure level of one of the more intense speech sounds. The values of β_0 and β_0' are given in Table XXX in Appendix 1. In general, for a listener with hearing loss β_H who is listening to speech, the threshold level $\beta_0'+\beta_H+4$ must replace the level β_0 in Eq. (37). Also for the case for calculating F the amplification α in Eq. (39) becomes α_F, given by Eq. (32). Consequently, when the system has this gain α_F the db above threshold Z_s of each critical band of received speech is given by

$$Z_s=B_s+\Delta B_s+R+\alpha_F+(\beta_t-68)+\kappa-(\beta_0'+\beta_H+4). \tag{40}$$

Let the value of Z_s for a flat response system be designated Z_F when the received speech is at the optimum for interpretation. For such a system

$$\bar{R}_1=\bar{R}_4=R=0$$

and

$$\alpha_0=\beta_H+12-\beta_t$$

and

$$\alpha_F=\alpha_0+68$$

and consequently

$$Z_F=B_s+\Delta B_s+\kappa-\beta_0'+8. \tag{41}$$

This is the equation from which the values in Fig. 11 were taken. The values of B_s, ΔB_s, κ and β_0' are given in Table XXX, Appendix 1. It is seen that Z_F is independent of both β_t and β_H since the amplification is always adjusted so that the received speech is at 68 db above the threshold of the listener. It gives the level above threshold of each critical band of speech when the received undistorted speech is at the optimum level for interpretation. The values of Z_F are also tabulated *versus* frequency in Table XXX of Appendix 1.

Substituting this value of Z_F and the value of α_F from Eq. (31), Eq. (40) becomes

$$Z_s=Z_F+R-\bar{R}_1+\gamma(\bar{R}_1-\bar{R}_4). \tag{42}$$

TABLE XVI. Constants for systems I, II, and III.

System	$\bar{R}_1$	$\bar{R}_4$	F_0	F_M	n_0	n_M	$\Sigma\Delta$	γ
I	0	−2.0	0.974	0.938	0.849	0.849	−0.06	0.006
II	−1.2	−2.8	0.987	0.987	0.870	0.870	−0.04	0.032
III	−2.5	−4.8	0.980	0.980	0.798	0.798	−0.05	0.181

$\alpha-\alpha_0=x_V+\phi\gamma(\bar{R}_1-\bar{R}_4)$

Let the point on the Z_s *versus* log f curve where the slope starts to exceed the critical slope be designated (Z_T,f_T). Also let R_T be the ordinate corresponding to f_T on the response curve R *versus* log f. Then the value of Z_s becomes Z_T when $R=R_T$ so that Eq. (42) reduces to

$$Z_T=Z_F+R_T-\bar{R}_1+\gamma(\bar{R}_1-\bar{R}_4). \tag{43}$$

If this is substituted in Eq. (38), the desired formula for calculating σ is obtained or

$$\sigma=75-\tfrac{1}{2}[Z_F+R_T-\bar{R}_1+\gamma(\bar{R}_1-\bar{R}_4)] \tag{44}$$

which is the slope in db per octave for the speech masking tangent line for the high frequency side. Let y be the ordinate either of the speech masking curve on the low frequency side or of the speech masking tangent constructed as just outlined. Then the increment

$$\Delta Z=y-Z_s \tag{45}$$

gives the number of db that the threshold level of any critical band of speech is raised due to the masking of one component of a speech sound by another component. Consequently, under these circumstances the critical value $\bar{R}-R_c$ becomes, instead of Eq. (35),

$$\bar{R}-R_c=Z_F+5+\Delta Z. \tag{46}$$

The work involved in constructing the Z_s *versus* log f curve to find the values of ΔZ can be avoided and ΔZ determined directly from the R *versus* log f curve by noting the following relationships.

If Z_s from Eq. (42) is substituted in Eq. (45), then

$$\Delta Z=y-Z_F-R-\gamma(\bar{R}_1-\bar{R}_4)+\bar{R}_1.$$

Now shift the Z_s *versus* log f curve and the corresponding masking curves downward by an amount $63+\gamma(\bar{R}_1-\bar{R}_4)-\bar{R}_1$ and let y' designate the ordinate of the shifted masking curves. Then

$$\Delta Z=y'-R+63-Z_F=\Delta R+63-Z_F \tag{47}$$

where ΔR is the difference between the ordinate of the shifted masking curves and the corresponding ordinate of the response curve. Consequently, the value of ΔZ can be determined from Eq. (47) if one knows how to construct the shifted masking curves on the R *versus* log f plot. Let y_T be the ordinate of the beginning point of such shifted masking curves for the frequency f_T. At this point $\Delta Z=0$ so from Eq. (47)

$$y_T=R_T-(63-Z_F). \tag{48}$$

So from 700 to 3000 c.p.s. where $Z_F=63$ db the ordinate

y_T is on the response curve and equal to R_T. At frequencies beyond these limits the starting point of the masking lines is below the response curve by the amount $63-Z_F$ at the frequency f_T. One can usually determine f_T from inspection of the R *versus* $\log f$ curve. It is the point where the response curve drops very suddenly as for a partial suppression filter system or at the frequency corresponding to the average ordinate over 180 cycles in the peak response for a resonant system.

11. CALCULATION OF THE F FACTOR

To calculate the F factor, one proceeds as follows. Examine slopes of the response curve under consideration. One can usually estimate whether the corresponding Z_s curve will or will not exceed the critical slope σ. If not, then no tangent curves need be drawn and Eq. (36) applies.

This is the case for many of the systems with which one deals. For filter systems and sharply resonant systems the value of σ is calculated and the tangent lines are drawn to determine ΔZ and from this by aid of Eq. (46) the value of $\bar{R}-R_c$ is calculated. Then by Eq. (34) the value of x_W is given by

$$x_W=68(\bar{R}-R)/[Z_F+5-(\Delta R+63-Z_F)]=r(\bar{R}-R) \qquad (49)$$

where the quantity r is defined by the equation

$$r=68/[Z_F+5-(\Delta R+63-Z_F)]. \qquad (49a)$$

Values of $W(x_W)$ for each value of x_W are obtained from Table XI. Then to calculate F one obtains the value of $\bar{R}$ from Eq. (33) and the values of D from Fig. 5. To obtain the integral indicated by Eq. (32) giving the value of F, one plots $\int_0^f Ddf$ as abscissa and $W(x_W)$ as ordinate. The area under such a curve is the required value of the integral. In general this area can be evaluated more simply by taking an average W at frequencies corresponding to equal intervals of $\int_0^f Ddf$.

In Table XII frequencies corresponding to 100 intervals of $\int_0^f Ddf$ are given. Usually the twenty frequencies shown in Table XII in italics are sufficient except near a sharp cut-off in the response curve, where for accuracy the one percent intervals should be taken.

This, then, gives the value of F designated F_M for the two-thirds of the speech sounds which involve masking of one component by another. Another calculation is made without masking curves, $\bar{R}-R_c$ being given by Eq. (35); the value so calculated is designated F_0. This value corresponds to the one-third of speech sounds which are in the first class. The value of F to use in Eq. (19) is, then,

$$F=\tfrac{2}{3}F_M+\tfrac{1}{3}F_0. \qquad (50)$$

12. DETERMINATION OF THE FUNCTIONS ϕ AND γ

We now return to the problem of calculating ϕ and γ, referred to at the end of Section 7. In Fig. 12 the articulation data shown by discrete points were obtained from tests of two telephone systems designated respectively as III–RN–1100–3.9 and III–HP–1500. The circles correspond to the resonant system and the crosses to the filter system. An attempt was made to choose a resonant system and a filter system having approximately the same value of F; but it is seen that the maximum articulation for the filter system is somewhat lower than that for the resonant system. Its

TABLE XVII. Derivation of articulation index A *vs.* effective gain x_V for an ideal flat response system.

(1) x_V	(2) $\alpha-\alpha_0$	(3) S_{23}	(4) A_{23}	(5) $\alpha-\alpha_0$	(6) S_3	(7) A_3	(8) $\alpha-\alpha_0$	(9) s_{3M}	(10) A_{3M}	(11) A_{23}/A_m	(12) A_3/A_m	(13) A_{3M}/A_m	(14) A Average	(15) A Adopted
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	5	0.078	0.053	5	0.016	0.035	5.1	0.183	0.049	0.054	0.036	0.051	0.047	0.05
10	10	0.180	0.121	10	0.057	0.087	10.2	0.362	0.106	0.123	0.090	0.112	0.108	0.11
15	15	0.318	0.221	15	0.142	0.149	15.3	0.544	0.188	0.225	0.160	0.198	0.194	0.20
20	20	0.472	0.323	20	0.305	0.257	20.3	0.698	0.286	0.330	0.268	0.301	0.300	0.30
25	25	0.615	0.418	25	0.491	0.365	25.4	0.822	0.411	0.427	0.380	0.433	0.413	0.41
30	30	0.731	0.521	30	0.685	0.505	30.4	0.897	0.542	0.532	0.525	0.570	0.542	0.54
35	35	0.817	0.614	35	0.820	0.637	35.4	0.935	0.655	0.627	0.664	0.690	0.660	0.66
40	40	0.880	0.707	40	0.892	0.736	40.4	0.953	0.730	0.722	0.766	0.767	0.752	0.75
45	45	0.922	0.783	45	0.927	0.802	45.4	0.967	0.810	0.800	0.836	0.853	0.830	0.83
50	50	0.949	0.847	50	0.949	0.858	50.4	0.972	0.85	0.863	0.894	0.895	0.884	0.89
55	55	0.964	0.905	55	0.961	0.900	55.4	0.977	0.90	0.924	0.940	0.947	0.937	0.94
60	60	0.971	0.955	60	0.967	0.940	60.4	0.980	0.94	0.975	0.980	0.990	0.982	0.98
65	65	0.975	0.980	65	0.969	0.955	65.4	0.981	0.95	1.000	0.995	1.000	0.998	1.00
70	70	0.975	0.980	70	0.970	0.960	70.4	0.981	0.95	1.000	1.000	1.000	1.000	1.00
75	75	0.973	0.970	75	0.970	0.960	75.4	0.980	0.940	0.990	1.000	0.990	0.993	0.99
80	80	0.970	0.950	80	0.970	0.960	80.4	0.979	0.925	0.970	1.000	0.974	0.981	0.98
85	85	0.966	0.920	85	0.970	0.960	85.4	0.977	0.900	0.940	1.000	0.947	0.962	0.96
90	90	0.961	0.890	90	0.970	0.960	90.4	0.973	0.860	0.910	1.000	0.905	0.938	0.93
95	95	0.954	0.863	95			95.4			0.882				0.90
100	100	0.949	0.847	100						0.865				0.87
105	105	0.941	0.825	105						0.842				0.85
110	110	0.934	0.808	110						0.825				0.83
115	115	0.927	0.793	115						0.810				0.81
120	120	0.917	0.775							0.792				0.79

maximum was the closest one for which articulation data were available. The solid line through the crosses is calculated for an ideal filter system having the same value of Fp, namely $Fp=0.704$, as for the resonant system.

For the ideal filter system the value of V is approximately equal to that for an ideal flat response system at the same level above threshold, in accordance with the statement included in the first paragraph of Section 7. In this case γ is nearly equal to zero so that, for the filter, Eq. (29) becomes

$$\alpha-\alpha_0=x_V$$

whereas for the resonant system, γ is not zero, hence

$$\alpha-\alpha_0=x_V+\phi\gamma(\bar{R}_1-\bar{R}_4).$$

Thus for these two systems the same value x_V of the effective gain corresponds to values of $\alpha-\alpha_0$ which are not equal.

The two systems which have been chosen were composed of parts that introduced no peculiar types of distortion. Moreover, we shall consider only values of $\alpha-\alpha_0$ which are less than 55 db. It follows that in Eq. (19) the factors E and H are each equal to unity, and therefore

$$Ap=VFp.$$

The two systems have been chosen so that the value of Fp is the same for each. It follows that for these two systems equal values of Ap correspond to equal values of V. Consequently, equal values of articulation correspond to equal values x_V of the effective gain.

Comparing the two curves in Fig. 12, two points having equal values of articulation are displaced by an amount $\Delta\alpha$ which is the difference between $\alpha-\alpha_0$ for the resonant system and $\alpha-\alpha_0$ for the filter system, for the same x_V. Hence

$$\Delta\alpha=\phi\gamma(\bar{R}_1-\bar{R}_4).$$

Values of $\phi\gamma(\bar{R}_1-\bar{R}_4)$ were obtained in this manner from the displacement $\Delta\alpha$. The value of $\bar{R}_1$ and of $\bar{R}_4$ were calculated by aid of Eqs. (23) and (26) from the response characteristic of the resonant system III–RN–1100–3.9 given in Fig. 30. It was found that $\bar{R}_1-\bar{R}_4=13.6$ db. Consequently,

$$\phi\gamma=\Delta\alpha/13.6.$$

The experimental values of $\Delta\alpha$ obtained from the curves in Fig. 12 are shown in Table XIII. If the maximum value of γ is taken as 1.25 for this resonant system, then the values of ϕ obtained from these data are given in the fourth row. It will be seen later that γ for this system is 1.25. After similar calculations with other resonant systems, the values of ϕ given in the last row were adopted.

As stated above, the factor γ is dependent upon the shape of the response curve. It is found to be more intimately related to the curve W *versus* $\int_0^f Ddf$, which is obtained from the response curve. In Fig. 13 are shown two such curves for systems III–RN–1100–39. and III–HP–1500, whose response curves are given in Figs. 30 and 26. It will be remembered that the areas under each of these curves give the corresponding values of F. Now if we define another quantity n by the equation

$$n=\int_0^\infty W'\cdot D\cdot df$$

where W' is any function of x_W such that W' is less than or equal to W for all values of x_W, then the ratio

$$x_\gamma=n/F=\int_0^\infty W'\cdot D\cdot df\bigg/\int_0^\infty W\cdot D\cdot df$$

is a variable which is closely correlated with the shape factor γ. For a flat system or a filter system $x_\gamma=1.0$ and for all other systems x_γ will be less than unity, being least for very resonant systems. After a choice of W' is made, x_γ can be related to γ by the experimental results on systems having a wide variety of response curve shapes.

About the most simple form of W' is $W'=1.0$ for $x_W\geqq 0$ and $W'=0$ for all other values of x_W. Then for system III–RN–1100–3.9 the value of n is equal to the area in Fig. 13 included under the curve between abscissas 0.20 and 0.43, while for system III–HP–1500 the value of n is almost equal to F. Thus the value of x_γ for the first system is $0.23/F=0.307$ and for the second system it is almost unity.

This method of calculating n was tried and the corresponding x_γ was related to γ and the resulting calculations gave a fairly good agreement with the observed results. After studying these results it was evident that a somewhat better choice for W' is

$$W'=10(W-0.9)$$

with the condition that $W'=0$ when $W\leqq 0.9$. So this value of W' was adopted. The value of n corresponding to this choice for system III–RN–110–3.9 is equal to 0.27 instead of 0.23, which is larger than for the previous choice of W' since the area in Fig. 13 corresponding to it includes the same area as before plus two small triangular areas on either side. For system III–HP–1500 the value of n is still approximately equal to unity.

Thus the value of x_γ will be given by

$$x_\gamma=10\int_{f_1}^{f_2}(W-.9)\cdot D\cdot df\bigg/\int_0^\infty W\cdot D\cdot df=n/F \quad (51)$$

where f_1 and f_2 are the two frequencies where $W=0.9$ so that the summation covers only the values of W greater than this value.

Since there will be two values of F and n when masking lines are necessary, the equation for calculating x_γ becomes

$$x_\gamma=\tfrac{1}{2}(n_0/F_0+n_M/F_M) \quad (52)$$

where n_0/F_0 is the value of Eq. (51) when the argument x_W of W is

$$68(\bar{R}-R)/Z_F+5$$

and n_M/F_M is the value of Eq. (51) when x_W is equal to

$$68(\bar{R}-R)/[Z_F+5-(\Delta R+63-Z_F)].$$

In the twenty-band method of calculation the process of getting n_0 and n_M is very simple. From each value of W greater than 0.9 one subtracts 0.9, and then takes one-half the sum to get n. One might have combined n_0/F_0 and n_M/F_M by some other method than by taking an arithmetical average but this simple method of combination was found satisfactory.

Then the value of γ can be related to x_γ by a functional relationship indicated by

$$\gamma=\psi(x_\gamma)$$

where the form of ψ is found from experiment. It was taken as zero for $x_\gamma=1.0$ and equal to 1.20 for small values of x_γ. The values between these limits were determined empirically from the articulation data and are shown in Table XIV. This relation was tried and found to give a fairly good fit for all the data but an even better fit is obtained by adding a small correction

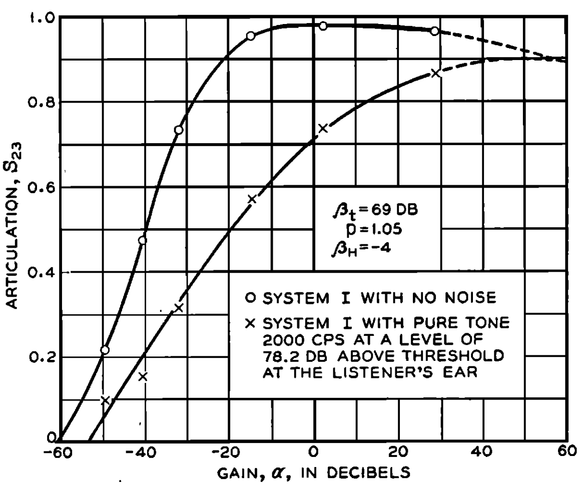


FIG. 17. Articulation of 1919–1920 high quality system without and with loud interference by a pure tone at frequency 2000 c.p.s.: The curves have been drawn to represent the points (observations).

term $x_\gamma\sum\Delta$ to ψ as follows.

$$\gamma=\psi(x_\gamma)+x_\gamma\sum\Delta. \quad (53)$$

It was found that high pass filter systems grow more rapidly to their maximum articulation as the gain is increased than do the low pass filter systems. This is what one would expect since the speech sounds are distributed over a smaller range of levels in the higher range of frequencies than at the lower range. For this reason the correction term $x_\gamma\sum\Delta$ is negative for the higher range of frequencies and positive for the lower range. For resonant systems it was found that this correction factor must be much smaller than for ideal filter systems. This is brought about by multiplying the correction term by x_γ since for ideal filters $x_\gamma=1$ and for resonant systems it is usually less than 0.4.

A number of methods of calculating this correction term were tried but none of them gave better results than the following simple one. Let Δ for each frequency be defined by the empirically determined values given in Table XV. Then

$$\sum\Delta=20\int_{f_1}^{f_2}\Delta\cdot D\cdot df+20\int_{f_3}^{f_4}\Delta\cdot D\cdot df, \text{ etc.},$$

where the limits are those corresponding to $W=0.99$. In the twenty-band method of computation this consists of adding the above Δ's together using only those bands where $W\geqq 0.99$. This sum is designated $\sum\Delta$.

13. DETERMINATION OF ARTICULATION DATA FOR AN IDEAL SYSTEM FROM DATA ON SYSTEMS I, II AND III

As stated in Section 8 these three relatively flat systems were not ideal, the response curves for them being given in Figs. 14–16. From these response curves the values of $\bar{R}_1$, $\bar{R}_4$, F_0, F_M, and γ were calculated and are given in Table XVI.

It is seen from the values in this table and from the formula at the bottom (see Eq. (29)) that the term $\phi\gamma(\bar{R}_1-\bar{R}_4)$ for systems I and II is always less than 0.1 db for all values of ϕ so that x_V and $\alpha-\alpha_0$ can be considered equal for these two systems. The values of x_V in steps of 5 db are written in the first column of Table XVII, which gives the steps taken in deriving the values of A *versus* x_V for the ideal system. The corresponding values of articulation S_{23} and S_3 were read from the experimental curves drawn through the observed points showing the relationship between $\alpha-\alpha_0$ and S_{23} or S_3 and are recorded in columns 3 and 6 of Table XVII. For system III the values of $\phi\gamma(\bar{R}_1-\bar{R}_4)$ are not negligible and are added to x_V to give $\alpha-\alpha_0$. The resulting values of $\alpha-\alpha_0$ are given in column 8. The corresponding observed articulation values of s_{3M} are given in column 9. The values of articulation index

TABLE XVIII. ΔB_s and K_m *vs.* f.

f	100	200	400	800	1000	2000	3000	4000
ΔB_s	7.5	7.8	9.3	10.9	11.5	13.7	15.1	15.5
K_m	9	11	14	13	12	7	−2	−11

TABLE XIX. Values of $J(x)$ *vs.* x.

x	0	1	2	3	4	5	6	7	8	9
0	1.0	1.0	1.0	0.99	0.99	0.98	0.98	0.97	0.96	0.95
10	0.94	0.92	0.90	0.88	0.85	0.82	0.79	0.76	0.72	0.68
20	0.63	0.58	0.53	0.48	0.44	0.40	0.36	0.33	0.30	0.27
30	0.24	0.21	0.18	0.15	0.12	0.10	0.08	0.06	0.04	0.02

TABLE XX. Values of a vs. β and f.

β	$f=100$	500	600	700	800	900	1000	1200	1400	1600	1800	2000	2500	3000	3500	4000
100	0.20	0.20	0.31	0.42	0.47	0.52	0.57	0.54	0.52	0.54	0.56	0.58	0.48	0.38	0.28	0.18
90	0.14	0.14	0.23	0.32	0.37	0.42	0.47	0.43	0.40	0.42	0.45	0.48	0.40	0.33	0.23	0.13
80	0.08	0.08	0.15	0.22	0.27	0.32	0.37	0.32	0.28	0.31	0.35	0.38	0.32	0.28	0.18	0.08
70	0.02	0.02	0.07	0.12	0.17	0.22	0.26	0.21	0.16	0.20	0.25	0.26	0.24	0.23	0.13	0.03
60	0	0	0.00	0.02	0.07	0.12	0.15	0.10	0.04	0.10	0.15	0.14	0.16	0.18	0.08	0.00
50	0	0	0	0	0.00	0.03	0.04	0.00	0	0.00	0.04	0.02	0.08	0.13	0.03	0.00
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0	0

for each value of articulation were found from Table III and tabulated in columns 4, 7, and 10. These values were divided by maximum articulation values for each system to obtain the values in columns 11–13. The average of these three values is given in column 14. The values in column 15, labeled "Adopted," were chosen as giving the best fit for all the data and correspond to the values of x_V in the first column. The articulation values given in Table X for an ideal system were determined from these articulation index values.

To show how well these values of A fit the observed data, the curves of articulation versus gain were calculated for systems I–III and are given in Figs. 14–16.

14. THE EFFECTS OF NOISE UPON ARTICULATION

In order to take account of the complicated action of an interfering noise in the reception of speech, the effects of the noise must be considered upon each of the factors in Eq. (19). These effects are generally expressed in terms either of the spectrum level B of the noise, or the masking M (threshold shift) which the noise causes in an ear of specified acuity. In the present section of this paper the relation of noise levels to masking will be considered. In the three following sections the effects of noise will be evaluated for the factors V, E, F, and H.

In our calculations here we require the value of M for an ear having the pure tone acuity level β_0 defined in Section 4, which corresponds to the typical articulation crew listener having the hearing loss $\beta_H=-4$ db.

Under the conditions of listening to speech in the presence of distributed noise, the expression for the level Z in db above threshold for each critical band of noise is no longer the same as Eq. (37). Instead, as stated in Section 10, the threshold $\beta_0'+\beta_H+4$ must replace β_0 so that

$$Z=B+\kappa-(\beta_0'+\beta_H+4) \tag{54}$$

where B is the spectrum level of the noise, κ the critical band width in db, and β_0' the threshold level for the critical bands of noise under the conditions of listening also to speech and for a listener with a hearing loss $\beta_H=-4$ db. It was seen that

$$\beta_0'=\beta_0+\Delta\beta_0 \tag{55}$$

where $\Delta\beta_0$ becomes as large as 8 db for low frequencies

TABLE XXI. Values of H vs. speed.

Speed	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7
$H=$	0.17	0.42	0.60	0.79	0.96	1.0	0.97	0.865	0.76	0.68	0.61	0.54	0.47

and is zero for the higher frequencies. It will be remembered that β_0 is the field intensity level of pure tones at the threshold of hearing for typical listeners in all the crews making articulation tests.

Experiments on masking have shown that M can be calculated from Z through the relationship

$$10^{M/10}=10^{Z/10}+1. \tag{56}$$

This satisfies the condition that when $B=-\infty$ (that is, no external noise present) then $M=0$, and also that Z and M are equal for levels of Z greater than about 20 db. This is subject to the limitation that the slope of the curve representing Z must not be greater than the critical slope σ (Eq. (38)) and it very seldom reaches such steep slopes. If it is greater, however, the same procedure as outlined for speech masking must be followed to get the real masking curve due to the noise.

Let B_R be the spectrum level at each frequency of the room noise at the listener's end and let h be the corresponding attenuation in decibels produced by holding the receiver cap to the ear. Then the spectrum level in the ear due to room noise leaking under the receiver

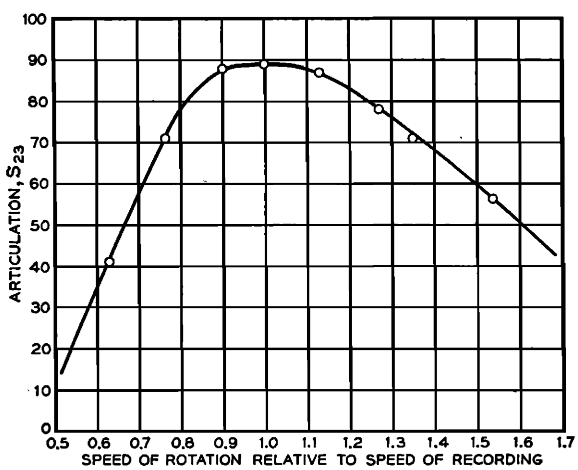


FIG. 18. Articulation vs. speed of rotation of a disk phonograph record of speech: Observations (points) and calculations (curve).

TABLE XXII. Values of H *vs.* frequency shift.

Frequency shift	−300	−200	−100	−50	−25	0	25	50	100	200	300	400	500
$H=$	0.56	0.66	0.77	0.85	0.94	1.00	0.98	0.96	0.89	0.78	0.69	0.58	0.49

TABLE XXIII. Values of H_R *vs.* reverberation time.

Reverberation time—seconds	0	1	2	3	4	5	6	7	8
$H_R=$	1.0	0.82	0.70	0.61	0.55	0.50	0.46	0.42	0.37

TABLE XXIV. Data for overloaded vacuum tube.

$\alpha_i=$	−40	−30	−20	−10	0	10	20	30
$\alpha_x=$	−17	−7	2	6	7.7	8.5	9	9
$\Delta\alpha=$	0	0	1	7	15.3	24.5	34	44
$S_{23}=$	0.79	0.79	0.79	0.78	0.745	0.65	0.45	

cap is B_R-h. Let R_s be the response at each frequency of the sidetone circuit of the telephone set. It gives the amplification or attenuation for the noise going through the microphone of the listener's subset to the ear of the listener and is measured for unity reproduction. For example, if R_s were zero db at all frequencies the noise would go through the microphone, the sidetone circuit and the receiver and arrive at the ear of the listener at the same level as that which would go directly into the ear of the listener with the receiver removed. Let B_L be the spectrum level of the line noise measured at the ear of the listener. If B is the spectrum level of the combined noises, then

$$10^{B/10}=10^{(B_R-h)/10}+10^{(B_R+R_s)/10}+10^{B_L/10}. \quad (57)$$

Having the spectrum level B at each frequency, then the value of Z at each frequency is given by Eq. (54) and M is obtained from Eq. (56). This holds for all levels except for critical slopes where M is greater than Z by an amount ΔZ. This ΔZ can be obtained in the same way as outlined for the masking of one speech sound by another but using Z_T for noise in Eq. (38) to obtain σ.

To obtain the masking M which speech suffers due to the presence of a pure tone or a combination of pure tones, when the components are separated by more than about 200 c.p.s., the method described above cannot be used. This masking has been obtained from two sets of data: (1) On the masking of one pure tone by another pure tone, and (2) on the masking of a narrow band of thermal noise by a pure tone. Since these data cover only a few levels and frequencies it was necessary to interpolate to obtain data corresponding to those tones used in articulation tests. This was done for both sets of data and the average taken for determining the masking. The two quantities thus averaged differ typically by about 0 to 12 db in the frequency regions of greater masking when the maskings are large, and by about 0 to 6 db when the maskings are small. The resulting $R-M$ curves are plotted in the upper plots of Figs. 35–37, which figures in the lower plots give the data showing the articulation results for various types of pure tone interference.

15. CALCULATION OF THE E AND V FACTORS—NOISE PRESENT

In Section 8 it was mentioned that when noise is present the factor E in Eq. (19) is regarded as the same as in the quiet—that is, equal to $E(\alpha-\alpha_0)$ where α_0 is the threshold gain when no noise is present.

The factor V is a known function of the effective gain x_V. To calculate the gain α which corresponds to a chosen value of x_V when noise with masking M is present, one proceeds in the same manner as when noise is absent except that the response at each frequency is taken as $R-M$ instead of R. The quantities $\langle\langle R-M\rangle\rangle_1$ and $\langle\langle R-M\rangle\rangle_4$ replace $\bar{R}_1$ and $\bar{R}_4$.* Therefore, the gain α_0 to reach the threshold level when noise is present is given by the following equation instead of Eq. (22):

$$\alpha_0=\beta_H+12-\beta_t-\langle\langle R-M\rangle\rangle_1. \quad (58)$$

Comparison of this equation with Eq. (22) shows that the threshold gain α_0 has been increased by $\bar{R}_1-\langle\langle R-M\rangle\rangle_1$ due to the presence of the noise. If the noise masking is constant with frequency (that is, equal to M db), then the threshold level is shifted M db due to the presence of noise.

It was seen that for $M>20$ db the value of M is equal to Z, hence by Eq. (54)

$$M=Z=B+\kappa-(\beta_0'+\beta_H+4).$$

For a noise having masking M constant with frequency, the value $\langle\langle R-M\rangle\rangle_1=\bar{R}_1-M$. Therefore, for such a noise

$$\alpha_0=(B+\kappa-\beta_0')-(\beta_t-8+\bar{R}_1)$$

and the gain to reach threshold is independent of the

TABLE XXV. Values of H *vs.* $\Delta\alpha$.

$\Delta\alpha$	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	35	40
H	0.997	0.994	0.998	0.976	0.963	0.950	0.93	0.91	0.89	0.85	0.81	0.76	0.72	0.68	0.63	0.52	0.40

* The angular bracket enclosing $R-M$ replaces the bar over R to indicate a weighted average as before. This is done throughout the text and the charts to facilitate the setting of type.

TABLE XXVI. Calculation of articulation for 3 cases of overloading. From response, $F=0.535$ (see Figs. 10 and 19). Let proficiency factor $p=1.0$.

					Case (1): No overloading $H=1.0$.					
$\alpha_i=$	0	5	10	15	20	25	30	40	50	60
$\Delta\alpha=$	0	0	0	0	0	0	0	0	0	0
$x_V=$	27.8	32.2	36.8	41.7	46.7	51.7	56.7	66.7	76.7	86.7
$x_E=$					69.5	74.5	79.5	89.5	99.5	109.5
$V=$	0.486	0.595	0.693	0.785	0.85	0.908	0.956	1.0	1.0	1.0
$E=$	1.000	1.0	1.0	1.0	1.0	0.992	0.980	0.935	0.87	0.83
$A=$	0.260	0.318	0.371	0.42	0.455	0.482	0.501	0.50	0.465	0.443
$S_3=$	0.312	0.413	0.500	0.578	0.625	0.658	0.682	0.680	0.639	0.610
					Case (2): Overloading starts at $\alpha_i=6$ db					
$\Delta\alpha=$	0	0	1	3.3	7	10.9	15.3	24.5	34	44
$x_V=$	27.8	32.2	35.8	38.4	39.7	40.8	41.4	42.2	42.7	42.7
$x_E=$					62.5	63.6	64.2	65	65.5	65.5
$V=$	0.486	0.595	0.673	0.721	0.747	0.756	0.772	0.785	0.793	0.793
$E=$	1.000	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$H=$	1.000	1.0	0.998	0.995	0.982	0.957	0.917	0.744	0.542	0.350
$A=$	0.260	0.318	0.358	0.384	0.394	0.391	0.378	0.313	0.228	0.149
$S_3=$	0.312	0.413	0.480	0.521	0.537	0.532	0.512	0.404	0.260	0.142
					Case (3): Overloading starts at $\alpha_i=26$ db					
$\Delta\alpha=$	0	0	0	0	0	0	1	7	15.3	24.5
$x_V=$	27.8	32.2	36.8	41.7	46.7	51.7	55.7	59.7	61.4	62.4
$x_E=$					69.5	74.5	78.5	82.5	84.2	85.0
$V=$	0.486	0.595	0.693	0.785	0.85	0.908	0.95	0.978	0.983	0.99
$E=$	1.000	1.0	1.0	1.0	1.0	0.992	0.98	0.97	0.964	0.96
$H=$	1.000	1.0	1.0	1.0	1.0	1.0	0.998	0.982	0.917	0.744
$A=$	0.269	0.318	0.371	0.420	0.455	0.482	0.492	0.487	0.475	0.391
$S_3=$	0.312	0.413	0.500	0.578	0.625	0.658	0.671	0.677	0.650	0.574

acuity β_H of the listener. The first term is the level above threshold of the critical bands of noise and is the same as the constant masking for a typical listener in the articulation crews. The second term is the level of the received speech above the unshifted threshold for the same listener. So in general for high levels of noise the threshold gain α_0 is determined by the talking level and the noise level and is approximately independent of the hearing acuity of the listeners unless the hearing loss is relatively great.

The value of x_V (see Eq. (29)) becomes

$$x_v=\alpha-\alpha_0-\phi\gamma[(\langle R-M\rangle)_1-(\langle R-M\rangle)_4] \qquad (59)$$

This is the effective gain x_V that determines $V(x_V)$.

The value of ϕ is the same as for the no-noise case. The value of γ is given by the same formula, namely, Eq. (53), but the values of x_γ and $\sum\Delta$ are different and are dependent upon the values of F_{NO} and F_{NM} now to be described.

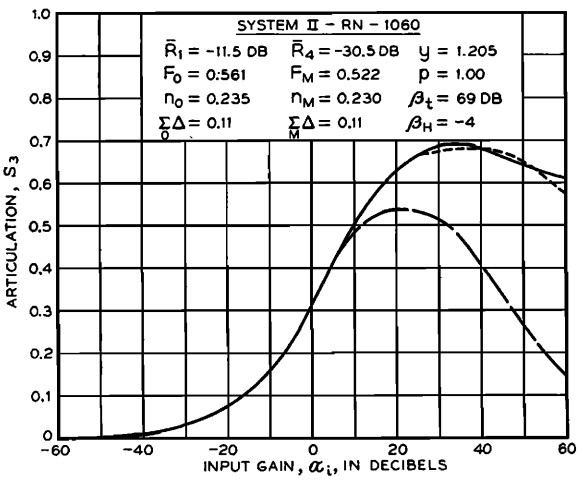


FIG. 19. Calculated articulation *vs.* gain for three conditions in regard to overload.

16. CALCULATION OF F FACTOR—NOISE PRESENT

There have been two points of view advanced as to how an observer interprets the speech sounds in the presence of a noise. The first point of view assumes that the relative position of the speech components with respect to the threshold in the noise determines the factor F in Eq. (19). According to this point of view the effective response has been lowered by the threshold shift M due to the noise, so that the quantity $R-M$ takes the place of R in determining the factor F. The second point of view, which was taken by one of the present authors in an earlier formulation of this theory, assumes that the level of the speech components with respect to each other is the principal influence in determining F. Then F is the same in the noise as in the quiet, except in so far as there is an increased masking of one speech component by another because of the higher received speech levels required in order to over-ride the noise.

The articulation tests indicate that some of the sounds of speech act in accordance with the first assumption, while the other sounds follow the second assumption. The sounds of the first class are those having components essentially in a single frequency region, constituting about one-third of the total number of fundamental speech sounds as previously noted in Section 10. For these sounds the quantity $R-M$ is used as though it were the actual response of a system without noise. Otherwise the calculation of the factor F proceeds as in the no-noise case except that for this first class of speech sounds there is no masking of one speech component by another, so that the quantity ΔZ due to speech masking described in Section 10 is considered to be equal to zero. Let the value of F calculated in this manner be designated as F_{NO}, the subscript N referring to the noise condition and the subscript O referring to the fact that there is no masking by speech components.

The second class of speech sounds consists of those sounds having components in more than one frequency region, which compose about two-thirds of the total number. For these sounds the calculation of the factor F uses the response R (not the difference $R-M$) and proceeds as in the case of no noise. However, the noise may have an effect upon this value of F because the slope σ of the speech masking tangent lines on the high frequency side will be greater than in the case of no noise, and may be greater than the critical slope. This is a consequence of the greater amplification α_F which must be used to reach the condition for which $V=1$. The quantity $R-M$ takes the place of R in determining α_F so that instead of Eq. (31), we have

$$\alpha_F=\alpha_0+68+\gamma[(\langle R-M\rangle)_1-(\langle R-M\rangle)_4]. \quad (60)$$

The increase in α_F is due chiefly to the increase in α_0,

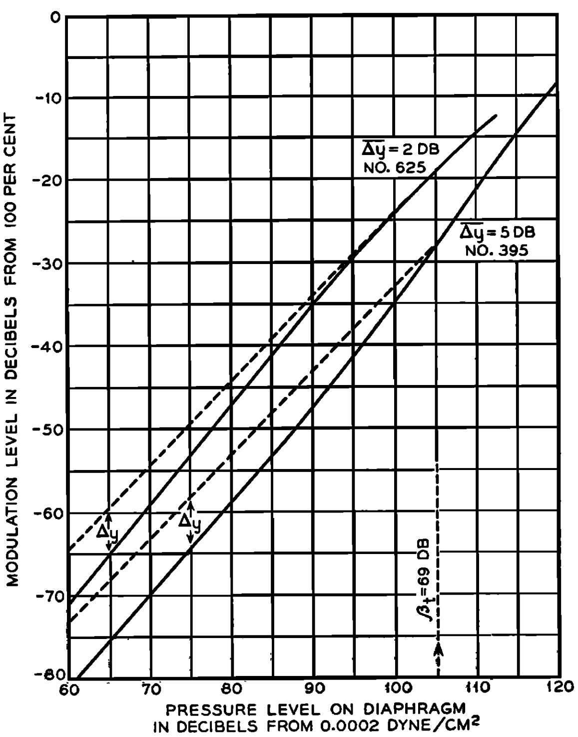


FIG. 20. Output *vs.* input characteristics of No. 395 and No. 625 carbon type microphones tested in a constant sound field at frequency 1000 c.p.s.

which was seen in Section 15 to be equal to M when M is constant with frequency.

The value of the critical slope σ_N when noise is present is given by the following equation instead of by Eq. (44):

$$\sigma_N=75-\tfrac{1}{2}\{Z_F+R_T-(\langle R-M\rangle)_1 +\gamma[(\langle R-M\rangle)_1-(\langle R-M\rangle)_4]\}. \quad (61)$$

The decrease in the slope as compared with σ for no noise is due principally to the term $(\langle R-M\rangle)_1$ being smaller than $\bar{R}_1$. The factor F_{NM} for this second class of speech sounds is then given by

$$F_{NM}=\int_0^\infty D\cdot W\left[\frac{68}{Z_F+5-\Delta Z}(\bar{R}-R)\right]df. \quad (62)$$

The factor F_{NO} for the other speech sounds is given by

$$F_{NO}=\int_0^\infty D\cdot W\left[\frac{68}{Z_F+5}[(\langle R-M\rangle)-(R-M)]\right]df \quad (63)$$

where $\langle R-M\rangle$ is given by

$$\langle R-M\rangle=\tfrac{1}{2}(\langle R-M\rangle)_1+\tfrac{1}{2}(\langle R-M\rangle)_4. \quad (64)$$

Then the final factor F_N to be used in Eq. (19) for the case when noise is present is given by

$$F_N=\tfrac{2}{3}F_{NM}+\tfrac{1}{3}F_{NO}. \quad (65)$$

The value of x_γ then becomes

$$x_\gamma=\tfrac{1}{2}[n_{NO}/F_{NO}+n_{NM}/F_{NM}] \quad (66)$$

where

$$n_{NO}=10\int_{f_1}^{f_2}\left\{D\cdot W\left[\frac{68[(\langle R-M\rangle)-(R-M)]}{Z_F+5}\right]-0.9\right\}df \quad (67)$$

$$n_{NM}=10\int_{0f}^{f_2}\left\{D\cdot W\left[\frac{68(\bar{R}-R)}{Z_F+5+\Delta Z'}\right]-0.9\right\}df \quad (68)$$

where $\Delta Z'$ refers to the effect of the masking tangent used when noise is present. The frequency limits used correspond to values of $W=0.9$.

The above may look like difficult calculations to make but it will be seen that when the chart method is used these calculations are very simple.

Similarly there will be two values of $\sum\Delta$, one obtained from the R *versus* f curve called $\sum_M$ and one obtained from the $R-M$ *versus* f curve called $\sum_0$. The value to be used in Eq. (53) for γ is

$$\sum\Delta=\frac{1}{2}\sum_M+\frac{1}{2}\sum_0. \quad (69)$$

17. CALCULATION OF H FACTOR DUE TO NOISE AND DETERMINATION OF K_m, α_m, a AND J

When the level of an interfering tone or noise becomes high there is an intermodulation between the speech sounds and the noise sounds so the factor H in Eq. (19) becomes less than unity. This is particularly apparent with pure tones of high level. For example, in Fig. 17 are shown the data for a 2000-cycle interfering tone at a level 78.2 db above threshold. The masking caused by this tone shifts the threshold for speech only a few decibels. The lowering of the articulation is caused chiefly by the intermodulation effect.

The deteriorating effect of the modulation would be expected to depend upon the intensity level β and the frequency f of the tone, and the average received total intensity level β_s of the speech. The latter is given by

$$10^{\beta_s/10}=\int_0^\infty 10^{(B_s+R+\alpha)/10}df. \tag{70}$$

For a flat response system where R and α are constant with frequency this reduces to

$$\beta_s=\beta_t+R+\alpha. \tag{70a}$$

For a system of variable response R it was considered that the approximate equation

$$\beta_s=\beta_t+\bar{R}_1+\alpha \tag{70b}$$

would apply adequately instead of the more general form of Eq. (70). The articulation data were not sufficiently accurate to show any difference between these two equations for β_s and the latter is much more simple to apply.

It would be expected that the maximum deteriorating effect of the modulation would occur when the intensity level β of the tone is approximately equal to some aspect of the speech closely related to the average received total intensity level β_s. Let the difference between β and β_s when this maximum deteriorating effect occurs be K_m. Then the amplification α_m when this occurs is given by

$$\alpha_m=\beta-\beta_t-\bar{R}_1-K_m. \tag{71}$$

The examination of the data for pure tones indicated that K_m at the various frequencies has the values shown in Table XVIII. The values of the peak factor in db ΔB_s for speech (see Appendix 1) are given in Table XVIII for comparison. The values are approximately equal in the range of frequencies below 1500 c.p.s. In this range K_m seems to be closely identified with the peak factor. For the higher range of frequencies some other effect is operating to make K much less than the peak factor.

The factor H was chosen of the form

$$H=1-aJ(\alpha_m-\alpha) \tag{72}$$

where the function J varies between zero and unity, a is constant dependent only upon the noise, and α_m has the same meaning as in Eq. (71). When $\alpha_m-\alpha=0$, then J is unity. When $\alpha_m-\alpha$ becomes large either positively or negatively, the value of J becomes zero. It was taken as symmetrical about α_m and approximately equal to zero when $\alpha_m-\alpha=40$ db so that $J(x)=J(-x)$. The values between zero and 40 were inferred from the articulation data. The values chosen are shown in Table XIX.

The value of α_m determines the amplification which gives the maximum deterioration but the amount of this deterioration is determined by the quantity designed a in Eq. (72). For a pure tone noise this quantity depends upon the intensity level β of the tone and also its frequency f, or

$$a=\varphi(\beta,f).$$

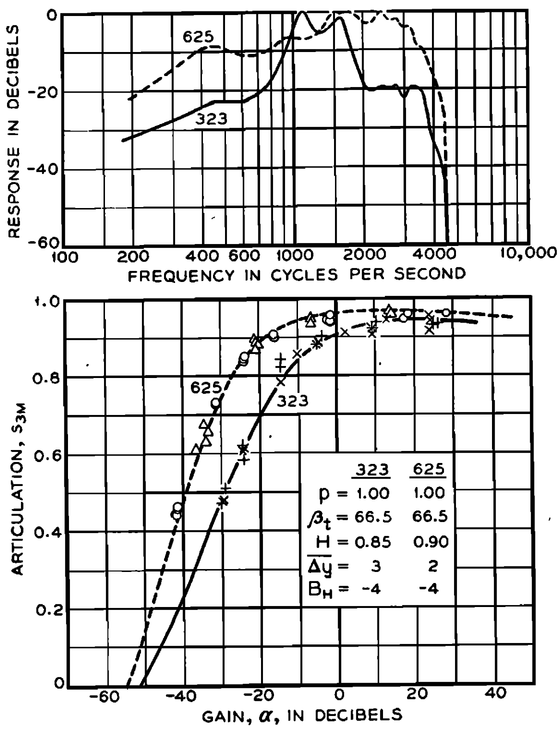
In the middle range of frequencies it was found that a was approximately proportional to $\beta-40$ but this relation did not hold at the high and low frequencies. So the function represented by the set of values given in Table XX was adopted as best representing the data.

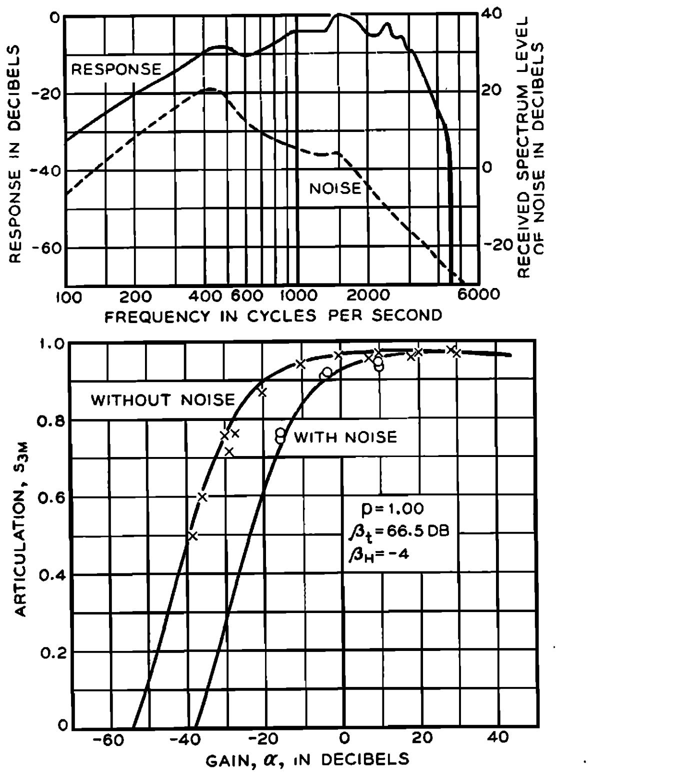
When more than one pure tone or a continuous noise is producing interference, the same procedure outlined above is used but we must now define the β and f for such noises. For these cases there are two compensating effects. The additional components produce more summation and difference tones but also they mask some of the regions which would otherwise let these tones be audible. This is also true of noises in general. The data seem to indicate that as the number of interfering tones of equal level increases, more of the interfering effect is taken over by the masking action and less by the H factor. It is rather complicated and difficult to formulate any simple rule for following the effect. Therefore, it was considered that the best approximation for a combination of tones was to take a value of the quantity a corresponding to the component which alone would give the greatest value of a and this was found to agree with the data as well as they are now known. The intensity level β corresponding to this component having frequency f is used in Eq. (71) to determine α_m and then a is found in Table XX corresponding to this β and f. For distributed noises the critical bands are considered as component tones. So the value of β is equal to $B+\kappa$ where B is the spectrum level of the noise, the particular value of $B+\kappa$ and f being that which gives the greatest value of a. This can be obtained from inspection of Table XX and a plot of $B+\kappa$. This value of a is used in Eq. (72) to find the factor H.

Using these relations for H fits all the available articulation data on systems with noise to within the accuracy of the data.

18. SPECIAL TYPES OF DISTORTION

Most of the systems for which calculations are shown involve no special types of distortion but only





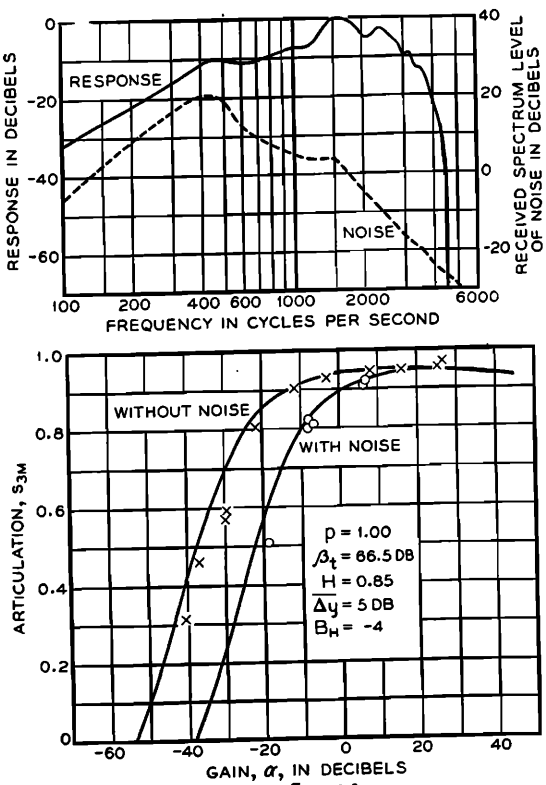


FIG. 21. Systems III–C–323, –625, containing carbon microphones: Response *vs.* frequency and articulation *vs.* gain.

FIG. 22. System III–C–395 containing carbon microphone, without and with a noise introduced electrically from a phonographic record of room noise.

FIG. 23. System III–HY–395 containing linear microphone (resembling No. 395 carbon microphone in certain properties), without and with noise.

non-uniform frequency response and noise. For other types of distortion it is difficult to make generalizations, for each problem has to be considered by itself. However, a method which gives results that are approximately correct is outlined here for common types of special distortion. The method uses the factor H for this purpose. If the distortion is in the transmitting or receiving mechanism and is independent of the amplification in the line, then the factor H can be determined from articulation tests using a known type of line. The value of H determined this way can then be used for any other type of line and the calculation can proceed as outlined above except that the factor F is replaced by $H \cdot F$. Four types of distortion which can be handled in this manner are now considered.

Reproducing Speed Different from Recording Speed

The first one is the distortion caused by multiplying all frequencies by a common factor such as is produced by a phonograph record of speech. The speed when reproducing is different from that while recording. A test of such a transmitting device indicates the factor H to be that given in Table XXI.†

For example, a system having a factor $pF=0.72$ and with amplification such that the speech is received at a level so that V and E are unity gave the results shown in Fig. 18. The curve was calculated for the various speeds. For any of these speeds an articulation *versus* gain curve can be made. For example, for speed equal to 0.8 normal $pF \cdot H=0.57$, which can be used as the factor F and calculations proceed as outlined above for no special type of distortion.

Frequency Shift

The second kind of distortion of this type is that due to shifting the frequency components such as may happen in carrier systems. Tests have shown that the factor H for such distortions is the value given in Table XXII. Although changes as much as 10 cycles are detectable, shifts as much as 200 produce only moderate distortions not greater than those due to a very resonant system.

Reverberation in Rooms at Sending End and at Receiving End

The third type of distortion considered here is that produced by reverberation in a room at either the transmitting end or the receiving end of the line or both. There are many variables in this problem and one can expect to get only an approximately correct answer by the method to be described, but this answer frequently is very useful in guiding engineering of systems for transmitting speech.

Even if one measured the response of such systems between each speaker and each listener, the procedure so far described still would not be valid because the prolonging of the speech in the room is not taken into account. So it seems better to take the response of the system as that obtained in a room with no reverberation—that is, with perfectly absorbing walls—and then use a factor H_R for rooms of various reverberations.

Since we are dealing with speech, this reverberation should be the average of that obtained with frequencies corresponding to the articulation importance function discussed above if the prolonging of the sound were not involved. Taking both effects into account, it is estimated that the average reverberation time taken at 500, 1000, and 2000 c.p.s. will yield the best time to take for a single value. In his book[19] Knudsen gives some results from which the values of H_R were calculated as given in Table XXIII. These values were obtained by the speaker talking directly to the listeners, who were at various positions in a room having a volume of 2096 cubic feet. The experimental articulation values were considered to be of the type S_{23} with proficiency factor $p=0.875$. In any telephone system there are two rooms to consider. Let H_L be the factor corresponding to the room in which the listener is placed, and H_S that corresponding to the room in which the speaker is placed. Although there are no data to guide one, it is roughly estimated that H_S approaches values H_R given in Table XXIII when the speaker's lips are about five feet away from the usual type of microphone. This value estimated to be five feet depends upon many factors but this figure is given for illustrative purposes. For smaller distances the value H_S to be used for microphones is increased. If one interpolates linearly from H_S to unity as the distance goes from five feet to very close talking, then the value H_S for the speaking end of the line is given approximately by

$$H_S=1-(1-H_R)d/5 \qquad (73)$$

where d is the distance in feet between the lips of the speaker to the position of the lips for close talking. Similarly for loudspeakers of the usual type one can go to about 20 feet (an estimate given for illustrative purposes) before reaching the value H_R so that for the listening end H_L is given by

$$H_L=1-(1-H_R)d/20. \qquad (74)$$

It must be emphasized that these relations for reverberant rooms are only very approximate since there are many factors not taken into account such as shape of the room, speed of talking, size of microphone and loudspeaker, etc., but this method will yield calculated results which show the general relation between reverberation in the speaking and listening rooms, the response, the volume of received speech, the noise

† Tables XXI–XXIV were derived from Figs. 145, 146, and 144 of reference 9.

[19] V. O. Knudsen, *Architectural Acoustics* (John Wiley and Sons, Inc., 1942).

conditions and the articulation obtained. Having obtained H_S and H_L, then the factor F is replaced by $F\cdot H_S\cdot H_L$ and the procedure is the same as for non-reverberant rooms or close talking and listening. Research work to obtain more accurate relationships of this kind is greatly needed.

Overloading

The next type of special distortion considered here is that due to overloading. It is common knowledge that overloading is very frequent in vacuum tube amplifiers. Let α_i be the input level to the vacuum tube amplifier, and α_x be the output. Then in Table XXIV is shown a typical set of articulation data when such an amplifier is in the system.

For input levels below −20 db the difference between α_x and α_i is constant and equal to 23 db, which is the gain of the amplifier before overloading starts. After overloading starts, this difference becomes progressively smaller and then negative as the input level increases. The quantity

$$\Delta\alpha=23-(\alpha_x-\alpha_i)$$

is a measure of the overloading and is given in the third row.

The response characteristics of the system from which the data in Table XXIV were obtained was not accurately known but from Table III it is evident that corresponding to $S_{23}=0.79$, $Ap=0.58$. Therefore, if the value of Ap corresponding to each value of S_{23} is divided by 0.58, one finds the value of H. Such a set of values of H *versus* $\Delta\alpha$ is given in Table XXV.

In dealing with such systems one must remember it is the output level which governs the received speech level and thus determines the effective gain. For illustration the calculation for system II–RN–1060, whose response curve is given in Fig. 10, will be made here. Three cases will be considered: (1) No overloading, (2) overloading starts at $\alpha_i=6$ db, and (3) overloading starts at $\alpha_i=26$ db. The values of $\bar{R}_1$, $\bar{R}_4$, α_0 and γ which are obtained from the response are given in Fig. 10. From these values one obtains for case (1)

$$x_V=\alpha_i+49.5-22.8\phi$$
$$x_E=\alpha_i+49.5.$$

For cases (2) and (3)

$$x_V=\alpha_i-\Delta\alpha+49.5-22.8\phi$$
$$x_E=\alpha_i-\Delta\alpha+49.5$$

but the values of $\Delta\alpha$ corresponding to α_i are different in cases (2) and (3). The calculations for the three cases are given in Table XXVI. The results are plotted in Fig. 19. An examination of this calculation shows that when overloading occurs, E becomes greater but both V and H become smaller. For case (2) the V and H factors are controlling and produce the lowering in articulation shown. But for case (3) these two effects approximately balance for the first 15-db overloading and then the H factor decreases faster than E increases.

19. CARBON MICROPHONE DISTORTION

In carbon microphones there are distortions not accounted for by the F factor as calculated. There are two effects which enter into the calculation in a different way. The first is due to the output *versus* input characteristics of the carbon microphones not being linear and thus the speech sounds are compressed or expanded on the intensity level scale.‡ The second is due to harmonic, summation, and difference tones being generated which affect the interpretation of the sounds.

In Fig. 20 are shown the output *versus* input characteristics for the No. 625 and No. 395 microphones. As indicated, these results were obtained for a single frequency of 1000 c.p.s. Very similar results will be obtained for any other frequency except those near a resonant peak. Since the loud sounds are 4 or 5 db above the average and also since for close talking the level is about 30 db above that at one meter which has been termed the "talking level," it follows that the level at the diaphragm of the microphone for the more intense sounds, due to a caller having a talking level of 69 db, is about 105 db as indicated in Fig. 20. This pressure level at the diaphragm varies with the type of mouthpiece, method of holding the microphone, etc. For purposes of this calculation the 105-db pressure level is sufficiently accurate to correspond to a talking level of 69 db.

Let (x_0,y_0) be the coordinates of the point on the output *versus* input curve corresponding to the talking level β_t of the caller. Then

$$x_0=105+(\beta_t-69). \tag{75}$$

When a weak sound is impressed upon the microphone, the input is dropped to a value x—that is, by an amount x_0-x. If the characteristics were linear the output would also drop by an amount x_0-x db. In general, however, the output drops by an amount $x_0-x+\Delta y$, as shown in Fig. 20. Consequently the low intensity sounds in the input are dropped to an even lower level in the output with respect to the loud sounds, so more of them will be masked and therefore F will be reduced. It is evident that this additional lowering has the same effect as reducing the response for these sounds. Since the speech sounds are distributed in level approximately uniformly over a 40-db range, an average value of Δy, designated $\langle\Delta y\rangle$, taken over the 40-db range below x_0, will be used. This $\langle\Delta y\rangle$ is subtracted from Z_F, the db above threshold of the speech sounds, in the formula for r in Section 11.

‡ A method of handling this effect in a computational manner was first suggested by P. V. Dimock and is given in the paper by French and Steinberg already cited in reference 3.

Calling this new value of r equal to r', then

$$r' = \frac{68}{Z_F + 5 - (\Delta Z + \langle \Delta y \rangle)} = r\frac{68}{68 - r\langle \Delta y \rangle} = r\left(1 + r\frac{\langle \Delta y \rangle}{68}\right). \quad (76)$$

In other words, this type of distortion has the effect of increasing the value of r by the factor $68/68-r\langle \Delta y \rangle$. For the No. 625 microphone the value of $\langle \Delta y \rangle$ is seen from Fig. 20 to be 2 db, and for the No. 395 microphone to be 5 db. No reliable data are available for the No. 323 microphone but it is estimated that $\langle \Delta y \rangle$ is equal to 3 db for this microphone.

If the output *versus* input current in Fig. 20 were above the broken line, then the speech sounds would be compressed and the value of r reduced instead of being increased. Such compression would help the interpretation if it could be accomplished without producing other deteriorating effects. Microphones have been made which approach the condition for $\langle \Delta y \rangle = 0$ but none has been made which compresses the range of speech sound levels.

In designing a microphone for commercial use many factors must be taken into consideration besides articulation under ideal conditions. For example, compressing the range of speech sound levels would raise the level of room noise going into the microphone through the sidetone circuit and into the receiver and thus interfere with the perception of the sounds being received. Expanding the level has the opposite effect.

The effect of producing harmonic, summation and difference tones is handled by making H less than unity. The value of H should be closely related to the level of the difference tones compared to the fundamental. Let two tones of frequency f_1 and f_1+200 c.p.s. and each at a pressure level of β be impressed upon the microphone diaphragm. Let the output level of the two fundamentals be β_1 db and β_1' db, respectively, and let the output level of the difference tone be β_d. Now let f_1 be changed progressively from 400 to 4000 c.p.s. and all of the output levels mentioned above measured. Call the average level of the two fundamentals $\bar{\beta}_1$, and of the difference tone $\bar{\beta}_d$. Then H should be related to $\bar{\beta}_1 - \bar{\beta}_d$. Since the effect upon H is small, it was considered that the relation could be made entirely dependent upon the level of the difference tone rather than considering the summation tones and the harmonics. Consequently the simple formula (77) was chosen:

$$H = 1 - 0.009[25 - (\bar{\beta}_1 - \bar{\beta}_d)]. \quad (77)$$

The constant 25 signifies that when the average level of the difference tone is down 25 db, H is unity. For larger values of $\bar{\beta}_1 - \bar{\beta}_d$ there is no effect upon articulation produced by such distortions and this equation must be so interpreted. The constant 0.009 was chosen to fit the data on the No. 395 and No. 625 microphones.

Consequently, for engineering purposes the value of H for the No. 395 and for the No. 625 microphones may be considered 0.85 and 0.9, respectively. No reliable data were available for the No. 323 microphone so the value of H was estimated to be the same as for the No. 395, namely, 0.85.

Using these values calculations were made upon systems having such microphones. In Fig. 21 are shown the results for the No. 625 and No. 323 microphone systems. The response curves are given at the top of the figure and observed and calculated data on articulation are given in the bottom part. It is seen that the calculated curves fit the observed points very well.

In Fig. 22 the results are given for a No. 395 carbon transmitter system for the cases when the listener is (1) in a quiet place, and (2) in a noisy place. The spectrum level of the noise at the listener's ear and the response curve for the system are given in the top part of the figure. The two articulation *versus* gain curves are given in the bottom part of the figure. A shift of 3 db in the calculated curves would give a better fit. It is considered that there is this much uncertainty in the combination of talking levels and response for these systems. Results were also taken on a system which had approximately the same response as for the No. 395 carbon microphone system but used a condenser transmitter. The results for this system are given in Fig. 23. The good agreement shown here and in the other curves is evidence that there is probably some error in the data of Fig. 22 but we were unable to find any.

20. DETAILED INSTRUCTIONS FOR USING THE CHART METHOD OF CALCULATION

The objective is to calculate a curve showing the relation between the gain α in the system and the articulation obtained by any speaker-listener pair. For this purpose charts have been prepared to make such calculations easier to follow. Charts 1 and 2 are used when no noise is present. Charts 3–5 are used when noise is present. Chart 6 contains all the tables used in the calculations except the one giving the relationship between articulation index and articulation. This is given in Chart 7.

Case I—No Noise and No Special Types of Distortion

1. Calculation of $\bar{R}_1$. (Use Chart 1.)

   1a. From the response curve of R *versus* f read the values of R corresponding to each of the frequencies in column 1 of Chart 1 and record in column 2.

   1b. From the proper tables read the values of the exponential $10^{R/10}$ and record in column 3.

   1c. Add the numbers in column 3 and take 1/20th of this sum. Then $\bar{R}_1 = 10 \log(1/20 \text{ sum})$.

2. Calculation of $\bar{R}_4$. (Use Chart 1.)

   2a. From the curve of R *versus* f read the values of R corresponding to each frequency in column 4 and record in column 5.

2b. Take the exponential indicated and record in column 6.
2c. Then $\bar{R}_4 = 40 \log(1/20 \text{ sum})$.

3. Calculation of $\bar{R}$. (Use Chart 1.)

3a. The value of $\bar{R}$ is the average of $\bar{R}_1$ and $\bar{R}_4$ and is recorded in the place indicated.

4. Determination of F_0.

4a. From the response curve read the values of R corresponding to each frequency in column 7 and record in column 9.
4b. Subtract each value of R from $\bar{R}$ and record in column 10.
4c. Multiply $\bar{R}-R$ by values of r found in column 8 and record the result in column 11.
4d. From these values find the corresponding values of W from Table I on Chart 6.
4e. The average of these twenty values of W is the desired value of F_0.

5. Determination of the value of n_0 and n_M. (Use Charts 1 and 2.)

5a. Subtract 0.9 from each value of W which is greater than 0.9.
5b. Add the resulting differences.
5c. One-half of the sum is the desired value of n_0.
5d. Perform a similar operation of the values of W on Chart 2 to find n_M.

6. Determination of $\sum\Delta$. (Use Chart 1.)
The values of Δ are given in the last column. Those values opposite values of W equal to unity or greater than 0.99 are added together to obtain $\sum\Delta$. The summation is recorded at the bottom of the chart at the place indicated.

7. Determination of F_M. (Use Chart 2.)
The procedure for calculating F_M is the same as for calculating F_0 except different values of r are used. The values of $\bar{R}_1$, $\bar{R}_4$, and $\sum\Delta$ are transferred from Chart 1 to Chart 2 to the places indicated.

7a. Determination of ΔR. (Use Chart 2.)
Construct the speech masking line on the high frequency side as follows. Determine the frequency f_T where the slope of the response curve exceeds the critical slope and record in the space indicated. This can usually be done with sufficient accuracy by inspection. From this frequency determine the corresponding response R_T from the curve and Z_T from values in column 2 of Chart 2. From this calculate y_T at f_T. The value of σ is obtained from the equation in the middle of the chart and so y at $2f_T$ can be calculated. A straight line is drawn between these points. Similarly on the low frequency side one follows the steps indicated on the chart to construct the speech masking curve on this side. The ordinates of these curves or series of straight lines corresponding to the frequencies in column 1 are recorded in column 3 under y. The values of R in column 4 are the same as on Chart 1. The values of ΔR are differences between y and R and are recorded in column 5.
7b. Calculation of r. (Use Chart 2.)
Determine the values of r from these values of ΔR by means of the formula in the middle of Chart 2 and record them in column 6. Whenever $\Delta R+62-F$ calculates to be zero or negative it is taken as zero. Also for large values of ΔR which make r negative, the value of r is taken as infinity.
7c. Transfer values of $\bar{R}-R$ from Chart 1 to Chart 2.
7d. Multiply these values by r to obtain $(\bar{R}-R)r$ and record in column 11.
7e. Find corresponding values of W and record in column 12.
7f. The value of F_M is the average of these values of W.

8. Calculation of γ. (Use Chart 2.)
From the values of n_0/F_0 and n_M/F_M one calculates the value of x_γ by the equation indicated in the middle of the chart. The value of ψ is read from Table (3), Chart 6. The value of γ is calculated by the equation indicated and recorded.

9. Calculation of α. (Use Chart 2.)
The table at the bottom of Chart 2 is used for this purpose.

9a. First calculate values of $\phi_\gamma(\bar{R}_1-\bar{R}_4)$ and record in the third row. The values of ϕ corresponding to each value of x_V are given in the second row.
9b. Calculate α_0 from the equation at the bottom of the chart and add to each value of x_V and record in the fourth row.
9c. Add values in the third and fourth row and record in the last row. These are the desired values of α. They are transferred to the last row of Chart 1.

10. Calculation of F and pF. (Use Chart 1.)
The value of p is determined from experimental data and given at the top of Chart 1. The value of F is obtained from F_0 and F_M by the formula in the middle of the chart. Its value is recorded. The value pF is then calculated and recorded in the place indicated.

11. Calculation of E. (Use Chart 1.)
The value of x_E is the db above threshold level for the listener in a quiet place or $\alpha-\alpha_0$. The values are recorded in the second row of the table at the bottom of the chart. Remember, no values below 70 db need be tabulated since E is unity for these values of x_E. The corresponding values of E are read from Table (2) of Chart 6.

12. Calculation of A and S. (Use Chart 1.)
The values in Chart 1 of pF, E and V are multiplied together to get Ap. These values of Ap are recorded in the fifth row opposite Ap in the chart. Corresponding articulation values are obtained from tables of Chart 7. These articulation values of S correspond to the α below them and consequently the desired curve of S *versus* α can be plotted and compared with experimental results.

Case II—Noise—No Special Types of Distortion

1. Calculation of the $R-M$ *versus* f curve. (Use Chart 3.)

1a. On Chart 3 write in column 4 the values of B, the spectrum level of noise.
1b. Add columns 3 and 4 and subtract β_H+4 to find values of Z and record in column 5.
1c. Calculate values of M from Z by the formula in the middle of Chart 3. It should be noticed that for Z greater than 20 db, $Z=M$ and consequently for such values no calculation is needed.

1d. Read the values of the response R for each frequency in column 1 from the graph showing R *versus* f for the system being calculated and record in column 7.
1e. Subtract the numbers in column 6 from those in column 7 and record in column 8.
1f. Plot the curve of $R-M$ *versus* f from the numbers in columns 8 and 1.

The last column is used later for determining a.

2. Calculation of $\langle\langle R-M\rangle\rangle_1$. (Use Chart 4.)

2a. From the curve of $R-M$ *versus* f found in step 1 read the values of $R-M$ corresponding to each of the frequencies in column 1 of Chart 4 and record in column 2.
2b. From the proper tables read the values of the exponential $10^{(R-M)/10}$ and record in column 3.
2c. Add the numbers in column 3 and take 1/20th of this sum. The value of $\langle\langle R-M\rangle\rangle_1=10\log(1/20 \text{ sum})$.

3. Calculation of $\langle\langle R-M\rangle\rangle_4$. (Use Chart 4.)

3a. Write down values of $R-M$ found from the curve of $R-M$ *versus* f determined in step 1 corresponding to each of the frequencies in column 4 and record in column 5.
3b. Take the exponential indicated and record in column 6.
3c. The value of $\langle\langle R-M\rangle\rangle_4$ is $\langle\langle R-M\rangle\rangle_4=40\log(1/20 \text{ sum})$.

4. Determination of $\langle R-M\rangle$. (Use Chart 4.)
This value is the average of $\langle\langle R-M\rangle\rangle_1$ and $\langle\langle R-M\rangle\rangle_4$ and is recorded in the place indicated.

5. Determination of F_{NO}. (Use Chart 4.)

5a. Again using the curve of $R-M$ *versus* f, write down in column 9 the values of $R-M$ corresponding to the frequencies in column 7.
5b. From the value of $\langle R-M\rangle$ obtained in step 4 subtract each of the values of $R-M$ in column 9 and record the resulting values in column 10. Opposite each negative value found in column 10 enter a value of W equal to 1 in column 12.
5c. Multiply each value in column 10 by r found in column 8 and record the resulting values in column 11.
5d. From each value found in column 11 find from Table (1) of Chart 6 the corresponding value of W and enter in the last column. Add the twenty values of W thus obtained and divide by 20 to find the desired value of F_{NO}.

6. Determination of F_{NM}. (Use Chart 5.)

6a. In column 2 record the values of R from the graph of R *versus* f.
6b. Take the exponential indicated and record in column 3.
6c. Add the twenty values together and divide by 20. The value of $\bar{R}_1$ is given by $\bar{R}_1=10\log(1/20 \text{ sum})$.
6d. Find the values of R corresponding to the frequencies in column 4 and record in column 5.
6e. Take the exponential indicated and record in column 6.
6f. Add the twenty values in column 6 together and divide by 20. The value of $\bar{R}_4=40\log(1/20 \text{ sum})$.
6g. The value of $\bar{R}$ is the average of $\bar{R}_1$ and $\bar{R}_4$.
6h. Determination of ΔR. (Use Chart 5.)
The speech masking curve is constructed the same as for the no-noise case except a different value of the critical slope is used. Its value σ_M is calculated by the formula at the top of the lower table in Chart 5. If y_M is the ordinate of the speech masking curve constructed as above, then $\Delta R=y_M-R$ and these values are recorded in column 9.
6i. Determination of r. (Use Chart 5.)
Determine r from the values of ΔR found in step 6*h* by the formula in the middle of Chart 5 and record in column 10. For values of $\Delta R+(63-Z_F)$ which are zero or negative, the value of $r=68/(Z_F+5)$. Also for large values of ΔR which make r negative, the value of r is taken as infinity. In other words, for such a high value of ΔR, the corresponding value of W is always zero.
6j. Subtract the values of R given in column 11 from the value $\bar{R}$ determined in 6g and record the values thus obtained in column 12.
6k. Multiply these values of $\bar{R}-R$ by r given in column 10 and record the resulting values in column 13.
6l. Look up the values of W corresponding to the values in column 13 from Table (1) of Chart 6 and record the results in column 14.
6m. The average of these twenty values of W is the value F_{NM} sought.

7. Determination of n_{NO}, n_{NM}, $\sum_{NO}$, and $\sum_{NM}$.
These values are determined respectively from Charts 4 and 5 in the manner described for the no-noise case.

8. Determination of γ. (Use Chart 5.)
From the values F_{NO}, F_{NM}, n_{NO}, and n_{NM} the value of x_γ is calculated by the equation indicated. From this value of x_γ the value of ψ is determined from Table (3) of Chart 6. The value of γ is then calculated by the formula indicated.

9. Calculation of the values of α. (Use Chart 5.)
The table at the bottom of Chart 5 is used for this purpose and the procedure is the same as outlined for the no-noise case. These values of α are now transferred to the last row of Chart 4.

10. Calculation of E. (Use Chart 4.)
This is the same as for the no-noise case and is given by the formula at the bottom of Chart 4.

11. Determination of F. (Use Chart 4.)
The value of F is given by

$$2F_{NM}+F_{NO}/3=F.$$

This is multiplied by the proficiency factor p and entered on Chart 4 at the place indicated.

12. Calculation of the values of H. (Use Chart 3.)

12a. Find from Chart 3 the value of $\beta=\kappa+B$ which yields the highest value of a when the pair of values of β and f are introduced into the table showing the values of a on Chart 6. For pure tones β is the intensity level and f the frequency. This maximum value of a is the value to use in the calculation and is entered at the top of Chart 3 in the place indicated. The values of β and f corresponding to this maximum value of a are the values β_m and f_m, which are also entered at the top of Chart 3 in the place indicated.
12b. From these values of β_m and f_m using the formula shown at the bottom of Chart 3 the value of α_m is obtained. The values of K_m are found from Table (6) of Chart 6.
12c. Enter in the first row of the table at the bottom of Chart 3 the values of α found on the last row of Chart 4.
12d. Enter the values of α_m in the second row.
12e. Enter the values of $\alpha-\alpha_m$ in the third row.
12f. Find the values of J corresponding to each of the values of $\alpha-\alpha_m$ from Table (4) of Chart 6.
12g. Multiply J by a and enter the result in the fourth row.
12h. The value of $1-J\cdot a$ is the value H sought and is entered in the sixth row. These values of H are now transferred to Chart 4 in the fifth row, opposite H.

13. Calculation of A and S. (Use Chart 4.)

The values in Chart 4 of pF, E, V, and H are multiplied together to get Ap. These values of Ap are recorded in the sixth row opposite Ap in Chart 4. The corresponding values of articulation are obtained from the Table in Chart 7. These values of articulation correspond to the α below them and consequently the desired curve of S *versus* α can be plotted and compared with experimental results.

CHART 1. Systems with no noise.

f	R	$10^{R/10}$	$\beta_t =$ f	R	$10^{R/40}$	f	r	R	$p = \bar{R} - R$	$(\bar{R}-R)r$	W	Δ
510			320			310	1.33					−0.03
640			500			470	1.06					−0.01
730			630			610	1.02					+0.01
820			750			740	1					+0.03
900			870			880	1					+0.04
980			990			1020	1					+0.04
1070			1120			1170	1					+0.03
1160			1260			1330	1					+0.02
1250			1410			1520	1					+0.00
1360			1570			1720	1					−0.02
1470			1760			1950	1					−0.04
1590			1960			2200	1					−0.05
1710			2190			2470	1					−0.06
1860			2440			2770	1					−0.06
2020			2740			3090	1					−0.05
2210			3090			3480	1.03					−0.04
2430			3520			3920	1.06					−0.02
2700			4080			4480	1.10					0.00
3100			4890			5210	1.17					+0.01
3930			6600			6370	1.39					+0.03

$SUM =$
$1/20(SUM) =$
$\bar{R}_1 =$
$\bar{R} = \frac{1}{2}(\bar{R}_1 + \bar{R}_4) =$

$SUM =$
$1/20(SUM) =$
$\bar{R}_4 =$

$n_0/F_0 =$
$pF =$

$SUM =$
$F_0 = 1/20(SUM) =$
$F_M =$
$F = \frac{2}{3}F_M + \frac{1}{3}F_0 =$

x_V	0	5	10	15	20	25	30	35	40	45	50	60	70	80	90	100	110	120
x_E																		
E																		
V	0.00	0.05	0.11	0.20	0.30	0.41	0.54	0.66	0.75	0.83	0.89	0.98	1.0	1.0	1.0	1.0	1.0	1.0
$A \cdot p$																		
S																		
c																		

$x_E = \alpha + \bar{R}_1 + \beta_t - 8 - (\beta_H + 4)$

$\Sigma\Delta =$ summation of all values of Δ where $W \geq 0.99 =$

$n_0 = \frac{1}{2} \sum_{W=0.9}^{W=1.0} (W - 0.9) =$

CHART 2. Systems with no noise.

f	Z_F	y	R	ΔR	r	$\bar{R}-R$	$(\bar{R}-R)r$	W
310	46							
470	59							
610	62							
740	63							
880	63							
1020	63							
1170	63							
1330	63							
1520	63							
1720	63							
1950	63							
2200	63							
2470	63							
2770	63							
3090	63							
3480	61							
3920	59							
4480	57							
5220	53							
6370	44							

$\bar{R}_1 =$ $\bar{R}_4$ $\Sigma\Delta =$

Construction of Speech Masking Curves

High frequency side $f_T =$

$R_T =$

$Z_T =$

$y_T = R_T - 63 + Z_F =$

y at $2f_T = y_T - \sigma =$

Low frequency side $f_T =$

$R_T =$

$y_T = R_T - 63 + Z_F =$

y at $0.7 f_T = y_T - 40 =$

y at $\frac{1}{2} f_T = y_T - 60 =$

y at $0.35 f_T = y_T - 70 =$

y at $\frac{1}{4} f_T = y_T - 75 =$

$\sigma = 75 - \frac{1}{2}[Z_F + R_T - \bar{R}_1 + \gamma(\bar{R}_1 - \bar{R}^4)] =$

$\Delta R = y - R$

$$r = \frac{68}{Z_F + 5 - (63 - Z_F + \Delta R)}$$

$n_M / F_M =$

$x_\gamma = \frac{1}{2}(n_0/F_0 + n_M/F_M) =$

$\psi =$

$\gamma = \psi + x_\gamma \Sigma\Delta =$

$SUM =$

$F_M = 1/20(SUM) =$

$x_V =$	0	5	10	15	20	25	30	35	40	45	50
$\phi =$	0.00	0.22	0.45	0.65	0.85	0.93	0.97	0.99	1.0	1.0	1.0
$\phi \cdot \gamma \cdot (\bar{R}_1 - \bar{R}_4) =$											
$\alpha_0 + x_V =$											
$\alpha =$											

$\alpha_0 = -\bar{R}_1 - \beta_t + 8 + (\beta_H + 4) =$

$\alpha = \alpha_0 + x_V + \gamma\phi(\bar{R}_1 - \bar{R}_4)$

$$n_M = \frac{1}{2} \sum_{W=0.9}^{W=1.0} (W - 0.9) =$$

21. COMPARISON OF OBSERVED AND CALCULATED ARTICULATION *VERSUS* GAIN CURVES FOR A LARGE VARIETY OF TELEPHONE SYSTEMS

Using the chart method described in Section 20, calculations of articulation were made upon a large variety of systems for which articulation data and response characteristics were known. Figures 14–16, and 21 to 46 inclusive show the results of such calculations. In general, one figure does not repeat the conditions considered in another, so that the total number of different telephone systems or of different sets of conditions computed is almost eighty. However, for purposes of comparison, one of the systems included in Fig. 28 is shown also in Fig. 29, and one system in Fig. 33 occurs also in Fig. 34.

The data necessary for the calculation are given on

CHART 3. Systems with noise.

f	β_0'	$\kappa-\beta_0'$	B	Z	M	R	$R-M$	$B+\kappa$
200	34	−16.8						
300	26	− 9.0						
400	20.2	− 3.2						
500	15.2	+ 1.8						
600	12.2	5.0						
700	9.9	7.4						
800	8.1	9.5						
900	6.8	11.0						
1000	5.6	12.4						
1250	3.7	14.8						
1500	2.4	16.6						
1750	1.7	17.7						
2000	0.6	19.3						
2500	−0.6	21.4						
3000	−1.2	22.7						
4000	−0.1	23.1						
5000	2.8	21.6						
6000	7.1	18.6						
7000	12.1	14.7						
8000	15.1	12.6						

$a_m =$

$(B+\kappa)m =$

$f_m =$

β_0' = threshold levels of critical band widths of speech,
κ = critical band widths in db for one ear listening,
B = spectrum level of noise,
$Z=B+(\kappa-\beta_0')-(\beta_H+4)$.

$$10^{M/10}=10^{Z/10}+1.$$

Calculation of H for noise.

$\beta=\kappa+B=$ $f_m=$ $a=$ $\alpha_m=$

α																	
α_m																	
$\alpha-\alpha_m$																	
$J(\alpha-\alpha_m)$																	
$a\cdot J$																	
$H=1-aJ$																	

$\alpha_m=\beta-\beta_t-\bar{R}_1-K_m.$

Chart 4. Systems with noise.

f	$R-M$	$10^{(R-M)/10}$	$\beta_t = f$	$R-M$	$10^{(R-M)/40}$	f	r	$R-M$	$p = \langle R-M\rangle^* - (R-M)$	$[\langle R-M\rangle - (R-M)]r$	W	Δ
510			320			310	1.33					−0.03
640			500			470	1.06					−0.01
730			630			610	1.02					+0.01
820			750			740	1					+0.03
900			870			880	1					+0.04
980			990			1020	1					+0.04
1070			1120			1170	1					+0.03
1160			1260			1330	1					+0.02
1250			1410			1520	1					0.00
1360			1570			1720	1					−0.02
1470			1760			1950	1					−0.04
1590			1960			2200	1					−0.05
1710			2190			2470	1					−0.06
1860			2440			2770	1					−0.06
2020			2740			3090	1					−0.05
2210			3090			3480	1.03					−0.04
2430			3520			3920	1.06					−0.02
2700			4080			4480	1.10					0.00
3100			4890			5220	1.17					+0.01
3930			6600			6370	1.39					+0.03

$SUM =$
$1/20(SUM) =$
$(\langle R-M\rangle)_1 =$
$\langle R-M\rangle = \frac{1}{2}(\langle R-M\rangle)_1 + \frac{1}{2}(\langle R-M\rangle)_4 =$

$SUM =$
$1/20(SUM) =$
$(\langle R-M\rangle)_4 =$

$n_{NO}/F_{NO} =$

$pF =$

$SUM =$
$F_{NO} = 1/20(SUM) =$
$F_{NM} =$
$F = \frac{2}{3}F_{NM} + \frac{1}{3}F_{NO} =$

x_V	0	5	10	15	20	25	30	35	40	45	50	60	70	80	90	100	110	120
x_E																		
E																		
V	0	0.05	0.11	0.20	0.30	0.41	0.54	0.66	0.75	0.83	0.89	0.98	1.0	1.0	1.0	1.0	1.0	1.0
H																		
$A \cdot p$																		
S																		
α																		

$x_E = \alpha + \overline{R}_1 + \beta_t - 8 - (\beta_H + 4)$

$\Sigma_0\Delta$ = summation of all values of Δ where $W \geq 0.99 =$

$n_{NO} = \frac{1}{3}\sum_{W=0.9}^{W=1.0}(W-0.9) =$

* See asterisk footnote on p. 120.

CHART 5. Systems with noise.

f	R	$10^{R/10}$	f	R	$10^{R/40}$	f	Z_F	ΔR	r	R	$\bar{R}-R$	$(\bar{R}-R)r$	W	Δ
510			320			310	46							−0.03
640			500			470	59							−0.01
730			630			610	62							+0.01
820			750			740	63							+0.03
900			870			880	63							+0.04
980			990			1020	63							+0.04
1070			1120			1170	63							+0.03
1160			1260			1330	63							+0.02
1250			1410			1520	63							0.00
1360			1570			1720	63							−0.02
1470			1760			1950	63							−0.04
1590			1960			2200	63							−0.05
1710			2190			2470	63							−0.06
1860			2440			2770	63							−0.06
2020			2740			3090	63							−0.05
2210			3090			3480	61							−0.04
2430			3520			3920	59							−0.02
2700			4080			4480	57							0.00
3100			4890			5220	53							+0.01
3930			6600			6370	44							+0.03

$SUM=$
$1/20SUM=$
$\bar{R}_1=$
$\bar{R}=\frac{1}{2}(\bar{R}_1+\bar{R}_4)=$
$r=\dfrac{68}{Z_F+5-(63-Z_F+\Delta R)}$

$SUM=$
$1/20SUM=$
$\bar{R}_4=$

$n_{NM}/F_{NM}=$
$x_\gamma=\frac{1}{2}(n_{NO}/F_{NO}+n_{NM}/F_{NM})=$
$\psi=$
$\gamma=\psi+\frac{1}{2}x_\gamma(\Sigma_0\Delta+\Sigma_M\Delta)=$

$SUM=$
$F_{NM}=1/20(SUM)=$

$\sigma_M=75-\frac{1}{2}\{Z_F+R_T-(\langle R-M\rangle)_1+\gamma[(\langle R-M\rangle)_1-(\langle R-M\rangle)_4]\}=$

$x_V=$	0	5	10	15	20	25	30	35	40	45	50
$\phi=$	0	0.22	0.45	0.65	0.85	0.93	0.97	0.99	1	1	1
$\phi\cdot\gamma[(\langle R-M\rangle)_1-(\langle R-M\rangle)_4]=$											
$\alpha_0+x_V=$											
$\alpha=$											

$\alpha_0=-(\langle R-M\rangle)_1+8-\beta_t+(\beta_H+4)=$

$\alpha=\alpha_0+x_V+\gamma\phi[(\langle R-M\rangle)_1-(\langle R-M\rangle)_4]$

$\Sigma_M\Delta=$summation of all values of Δ where $W\geq 0.99=$

$$n_{NM}=\frac{1}{2}\sum_{W=0}^{W=1.0}(W-0.9)=$$

each figure. The reader is referred to Table I of this paper for an explanation of the designations of these telephone systems and of the symbols used to represent the articulation, and to Section 3 for a description of the type of response. The quantities p, β_t and β_H have been explained in Sections 2, 5, and 7, respectively. Unless interference is indicated by notations on the plots, the tests were made under quiet conditions. For interfering pure tones the frequency and the level above threshold (L.A.T.) are indicated, and the difference $(R-M)$ is plotted *versus* frequency—where R=response and M=shift of threshold for speech sounds, caused by the tone. When there was a noise having energy distributed through a frequency interval,

CHART 6.

(1) Values of $W(x)$ vs. x

x	0	1	2	3	4	5	6	7	8	9	Difference
0	1.0	0.997	0.994	0.990	0.985	0.980	0.973	0.966	0.958	0.950	0.007
10	0.940	0.930	0.920	0.910	0.899	0.887	0.874	0.860	0.846	0.832	0.012
20	0.818	0.804	0.789	0.774	0.759	0.744	0.728	0.712	0.695	0.678	0.017
30	0.660	0.642	0.623	0.603	0.582	0.561	0.539	0.516	0.492	0.467	0.022
40	0.441	0.415	0.390	0.365	0.340	0.315	0.291	0.267	0.244	0.222	0.022
50	0.202	0.183	0.165	0.148	0.132	0.118	0.104	0.091	0.080	0.070	0.013
60	0.060	0.050	0.040	0.030	0.022	0.015	0.010	0.005	0.000		0.010

(2) Values of $E(x)$ vs. x

x	0	1	2	3	4	5	6	7	8	9
60	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
70	1.00	0.998	0.996	0.994	0.992	0.990	0.988	0.986	0.984	0.982
80	0.980	0.977	0.973	0.969	0.965	0.960	0.955	0.949	0.943	0.937
90	0.931	0.925	0.919	0.913	0.907	0.900	0.894	0.888	0.882	0.876
100	0.870	0.866	0.862	0.858	0.854	0.850	0.845	0.841	0.838	0.834
110	0.830	0.826	0.822	0.818	0.814	0.810	0.806	0.802	0.798	0.794
120	0.790	0.786	0.782	0.778	0.774	0.77	0.766	0.762	0.758	0.754

(3) Values of ψ vs. x_γ

x_γ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.3	1.20	1.20	1.20	1.20	1.20	1.20	1.198	1.196	1.194	1.192
0.4	1.185	1.176	1.167	1.158	1.149	1.140	1.128	1.114	1.097	1.083
0.5	1.070	1.060	1.045	1.030	1.010	0.990	0.970	0.940	0.905	0.870
0.6	0.840	0.805	0.770	0.740	0.710	0.685	0.655	0.625	0.595	0.560
0.7	0.530	0.500	0.475	0.445	0.415	0.390	0.360	0.335	0.310	0.285
0.8	0.265	0.240	0.215	0.185	0.155	0.130	0.110	0.090	0.070	0.055
0.9	0.045	0.035	0.025	0.015	0.008	0.003	0	0	0	0

(4) Values of $J(x) = J(-x)$

x	0	1	2	3	4	5	6	7	8	9
0	1.0	1.0	1.0	0.99	0.99	0.98	0.98	0.97	0.96	0.95
10	0.94	0.92	0.90	0.88	0.85	0.82	0.79	0.76	0.72	0.68
20	0.63	0.58	0.53	0.48	0.44	0.40	0.36	0.33	0.30	0.27
30	0.24	0.21	0.18	0.15	0.12	0.10	0.08	0.06	0.04	0.02

(5) Values of a in terms of β and f

β	$f=100$	500	600	700	800	900	1000	1200	1400	1600	1800	2000	2500	3000	3500	4000
100	0.20	0.20	0.31	0.42	0.47	0.52	0.57	0.54	0.52	0.54	0.56	0.58	0.48	0.38	0.28	0.18
90	0.14	0.14	0.23	0.32	0.37	0.42	0.47	0.43	0.40	0.42	0.45	0.48	0.40	0.33	0.23	0.13
80	0.08	0.08	0.15	0.22	0.27	0.32	0.37	0.32	0.28	0.31	0.35	0.38	0.32	0.28	0.18	0.08
70	0.02	0.02	0.07	0.12	0.17	0.22	0.26	0.21	0.16	0.20	0.25	0.26	0.24	0.23	0.13	0.03
60	0.00	0.00	0.00	0.02	0.07	0.12	0.15	0.10	0.04	0.10	0.15	0.14	0.16	0.18	0.08	0.00
50	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.00	0.00	0.00	0.04	0.02	0.08	0.13	0.03	0.00
40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00

(6) Values of K_m in terms of f_m

f_m	=	100	200	300	400	500	700	1000	1400	2000	3000	4000
K_m	=	9	11	13	14	15	14	12	10	7	−2	−11

the upper plot in the figure shows at each frequency the spectrum level of the noise—that is, the level of the intensity per cycle, in decibels from 10^{-16} watt/cm².

The remarkable agreement between observed articulations (shown by points) and calculated articulations (shown by curves) gives one considerable confidence that this method of calculation is valid for almost any kind of system that need be considered. The attention of the reader is directed to the instance of poorest agreement, which is shown in Fig. 29 by one of the observations (a circle point) displaced almost 8 db from the corresponding calculated curve. In this instance one cannot suppress the wish that more observations had been made at the lowest level. On the other hand, the reader should note the variety of noiseless systems which show a close agreement; also the systems with noise which are computed well, especially Fig. 36 where the effects are very different for the three different interfering tones, and Figs. 40 to 42 and 46 where each combination of restricted frequency band and intense noise causes a heavy but not uniform penalty.

22. APPLICATIONS

In this section calculations will be made for a variety of systems which are of general interest.

There is first considered a commercial telephone system consisting of one-mile No. 24 gauge cable with 9-db trunk into which is connected a filter cutting off all frequencies above 4000 c.p.s. The articulation obtained using such a system has been calculated in Fig. 47 for certain conditions involving no noise, and in Fig. 48 for a variety of conditions involving noise. The response of the system is shown in Fig. 48 by the curve R.

CHART 7. Articulation *vs.* articulation index.

A_p	s_{3M}	S_3	S_{23}	S_2	A_p	s_{3M}	S_3	S_{23}	S_2
0.01	0.041	0.004	0.014	0.010	0.51	0.882	0.693	0.721	0.794
0.02	0.080	0.008	0.028	0.023	0.52	0.887	0.704	0.732	0.802
0.03	0.118	0.013	0.043	0.040	0.53	0.891	0.714	0.742	0.810
0.04	0.154	0.018	0.058	0.058	0.54	0.896	0.725	0.752	0.818
0.05	0.189	0.024	0.073	0.075	0.55	0.900	0.736	0.761	0.825
0.06	0.222	0.031	0.088	0.093	0.56	0.904	0.746	0.771	0.832
0.07	0.254	0.040	0.103	0.111	0.57	0.908	0.757	0.780	0.839
0.08	0.285	0.049	0.118	0.129	0.58	0.912	0.767	0.790	0.846
0.09	0.314	0.060	0.133	0.146	0.59	0.916	0.777	0.799	0.852
0.10	0.342	0.071	0.148	0.164	0.60	0.919	0.787	0.807	0.858
0.11	0.369	0.084	0.163	0.182	0.61	0.922	0.797	0.815	0.864
0.12	0.395	0.098	0.177	0.199	0.62	0.925	0.806	0.822	0.869
0.13	0.421	0.112	0.191	0.217	0.63	0.928	0.815	0.829	0.874
0.14	0.444	0.127	0.205	0.234	0.64	0.931	0.824	0.836	0.879
0.15	0.466	0.144	0.219	0.252	0.65	0.934	0.832	0.843	0.884
0.16	0.488	0.156	0.233	0.269	0.66	0.937	0.839	0.850	0.889
0.17	0.509	0.173	0.247	0.286	0.67	0.939	0.846	0.857	0.894
0.18	0.529	0.186	0.261	0.304	0.68	0.942	0.854	0.864	0.898
0.19	0.548	0.201	0.275	0.323	0.69	0.944	0.861	0.870	0.902
0.20	0.567	0.216	0.289	0.342	0.70	0.947	0.868	0.877	0.906
0.21	0.585	0.231	0.303	0.362	0.71	0.949	0.875	0.884	0.910
0.22	0.602	0.247	0.318	0.382	0.72	0.951	0.882	0.890	0.914
0.23	0.618	0.263	0.333	0.402	0.73	0.953	0.888	0.895	0.917
0.24	0.634	0.279	0.348	0.422	0.74	0.955	0.894	0.900	0.920
0.25	0.649	0.295	0.363	0.441	0.75	0.957	0.900	0.905	0.923
0.26	0.663	0.312	0.378	0.460	0.76	0.958	0.906	0.910	0.926
0.27	0.677	0.329	0.393	0.478	0.77	0.960	0.911	0.915	0.929
0.28	0.690	0.347	0.407	0.496	0.78	0.962	0.916	0.921	0.932
0.29	0.703	0.365	0.422	0.514	0.79	0.963	0.921	0.926	0.934
0.30	0.715	0.382	0.437	0.531	0.80	0.965	0.926	0.931	0.937
0.31	0.727	0.399	0.452	0.548	0.81	0.967	0.930	0.935	0.939
0.32	0.738	0.416	0.467	0.564	0.82	0.968	0.935	0.939	0.941
0.33	0.749	0.433	0.482	0.580	0.83	0.969	0.939	0.943	0.943
0.34	0.759	0.450	0.497	0.596	0.84	0.970	0.943	0.947	0.945
0.35	0.769	0.467	0.512	0.611	0.85	0.972	0.946	0.950	0.947
0.36	0.778	0.483	0.527	0.625	0.86	0.973	0.950	0.953	0.949
0.37	0.787	0.499	0.542	0.639	0.87	0.974	0.953	0.956	0.951
0.38	0.796	0.515	0.558	0.653	0.88	0.975	0.956	0.959	0.952
0.39	0.805	0.531	0.574	0.666	0.89	0.976	0.959	0.961	0.954
0.40	0.813	0.547	0.589	0.679	0.90	0.977	0.961	0.963	0.955
0.41	0.821	0.563	0.604	0.692	0.91	0.978	0.963	0.965	0.956
0.42	0.828	0.578	0.617	0.704	0.92	0.979	0.964	0.966	0.958
0.43	0.835	0.592	0.630	0.715	0.93	0.980	0.965	0.967	0.960
0.44	0.842	0.606	0.642	0.726	0.94	0.980	0.967	0.969	0.961
0.45	0.848	0.619	0.654	0.737	0.95	0.981	0.968	0.970	0.963
0.46	0.854	0.632	0.665	0.747	0.96	0.982	0.970	0.972	0.964
0.47	0.860	0.644	0.677	0.757	0.97	0.983	0.971	0.973	0.966
0.48	0.866	0.657	0.689	0.767	0.98	0.983	0.973	0.975	0.967
0.49	0.871	0.669	0.700	0.776	0.99	0.984	0.974	0.976	0.968
0.50	0.877	0.681	0.711	0.785	1.00	0.985	0.976	0.978	0.970

It was obtained by W. Koenig of the Bell Telephone Laboratories. A proficiency factor of 0.9 was taken for a typical talker-listener pair. The heavy curve in Fig. 47 is for a talking level of $\beta_t=68$, which is considered an average conversational level, and for a listener having zero hearing loss $\beta_H=0$ db at all frequencies on a standard audiometer and for no room or line noise at the listener's ear.

Measurements upon a large number of speakers who were talking in a conversational manner over an experimental telephone system indicated that the average talking level of the different speakers varied from 56 to 74 db, or a range of 18 db, when the softest 5 percent and loudest 5 percent were excluded. Similarly, excluding the extreme 5 percent having poorest hearing and 5 percent having most acute hearing, the acuity level of a typical set of listeners varies over a range of 32 db from cases having a hearing loss of 20 db to cases having an acuity 12 db better than the average zero on the audiometer. Consequently, for the poorest listener

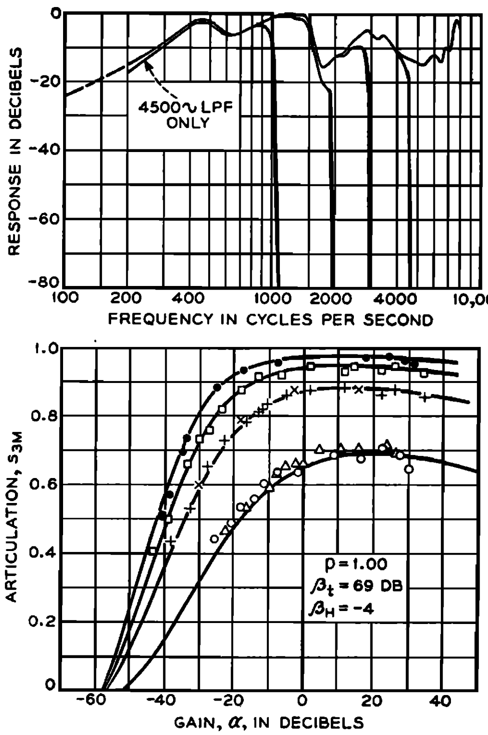


Fig. 24

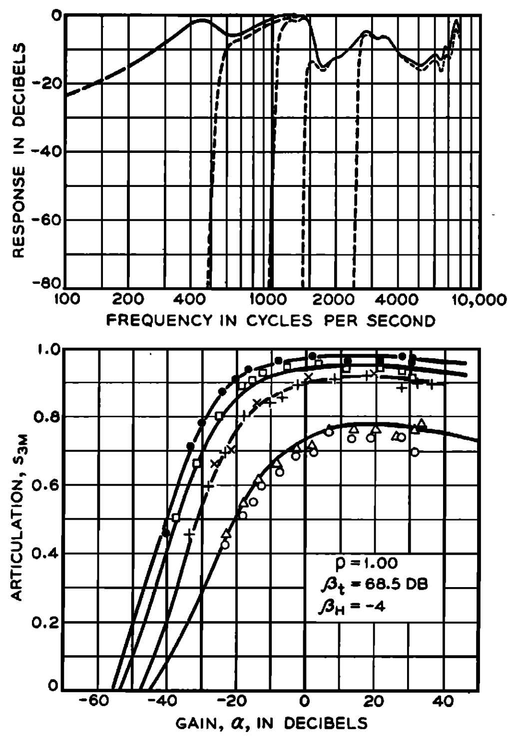


Fig. 26

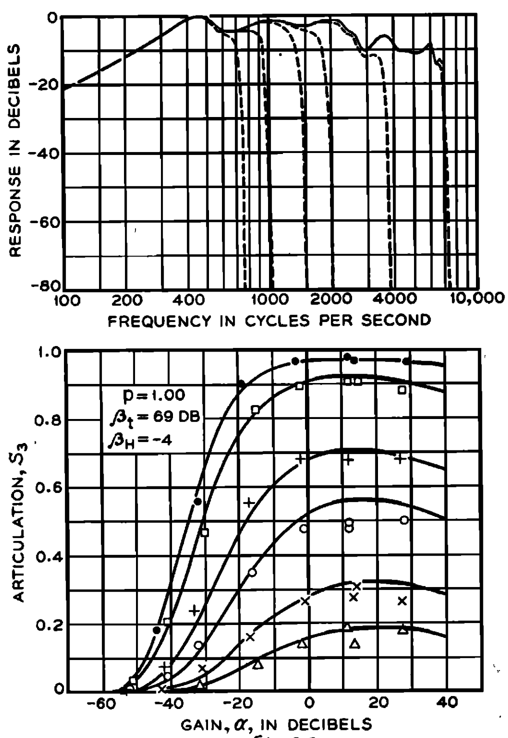


Fig. 25

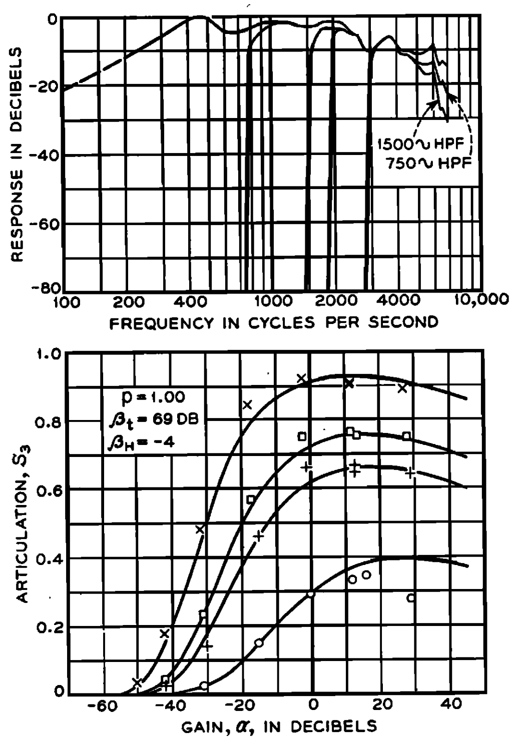


Fig. 27

FIG. 24. Filter systems III-LP-1000, −1850, −2850, −4500.
FIG. 25. Filter systems II-LP-750, −1000, −1500, −1950, −3700, −7000.
FIG. 26. Filter systems III-HP-500, −1000, −1500, −2500.
FIG. 27. Filter systems II-HP-750, −1500, −1900, −2850.

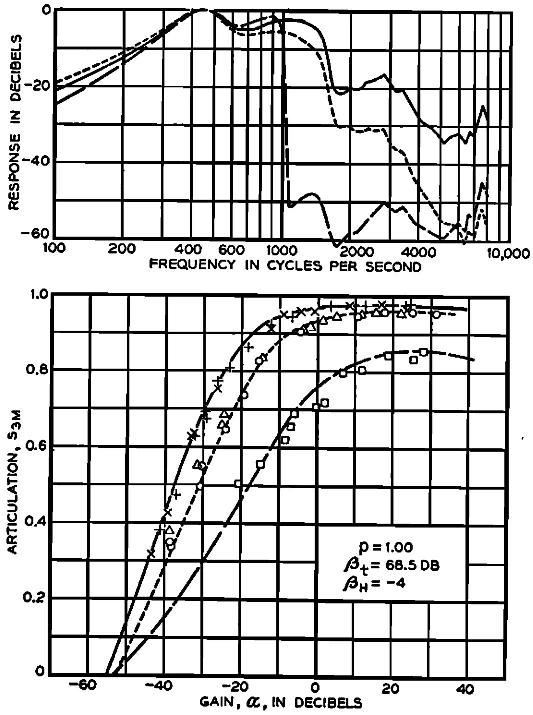


Fig. 28

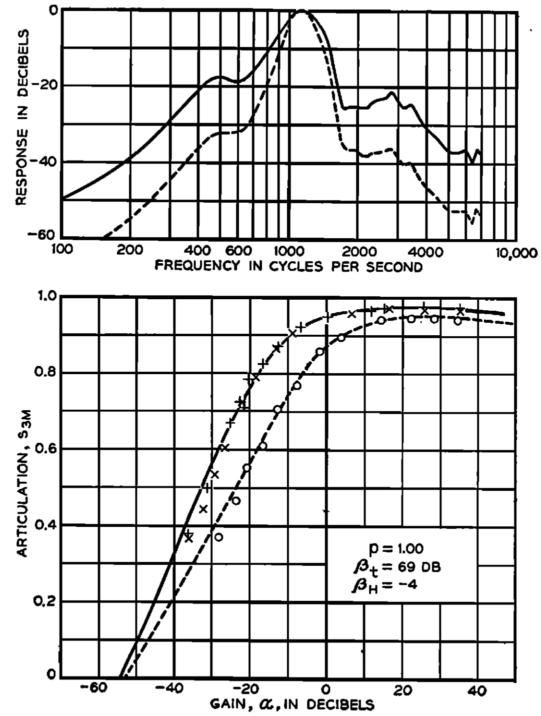


Fig. 30

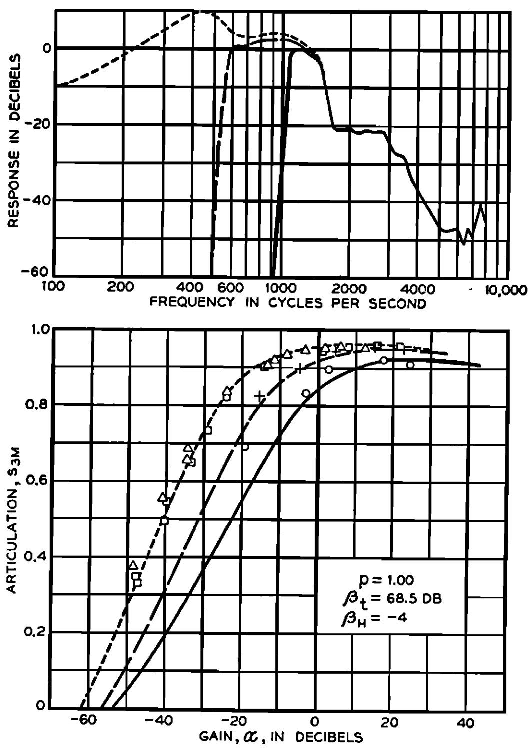


Fig. 29

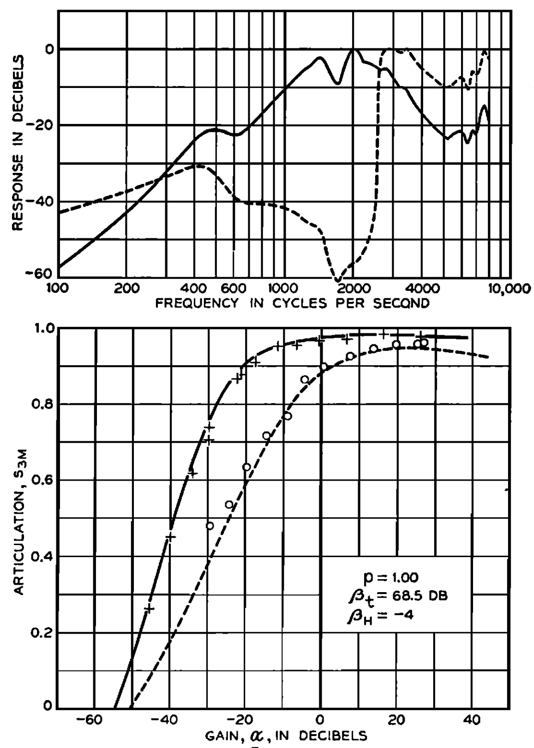
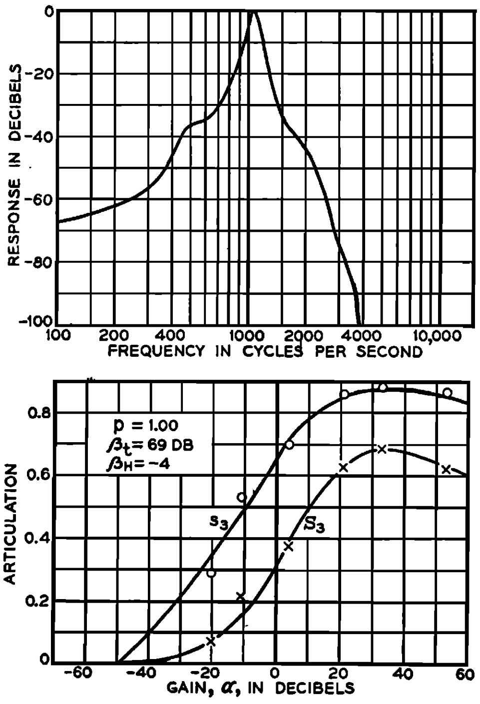
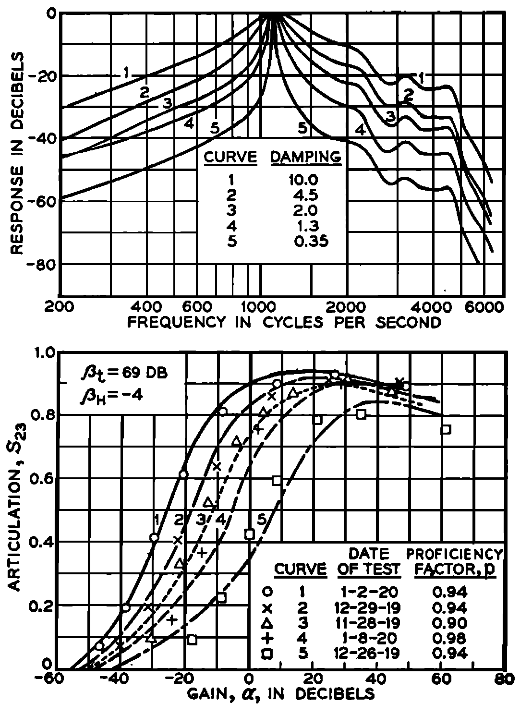
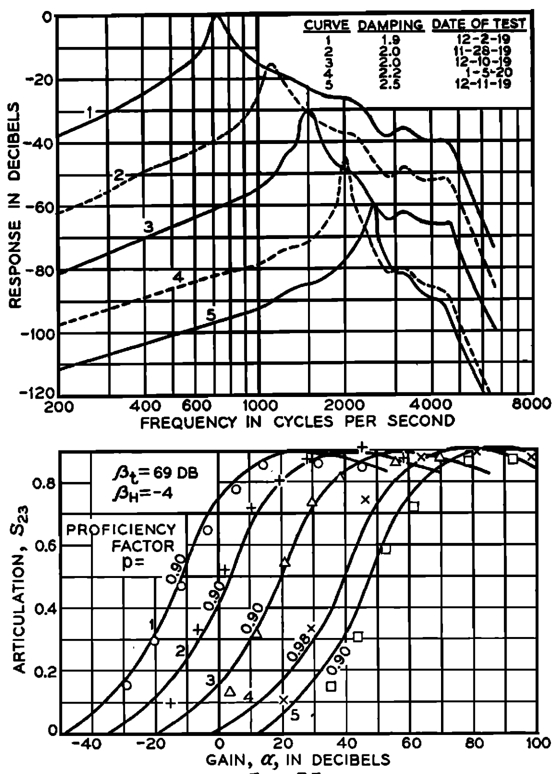


Fig. 31

FIG. 28. Rising loss systems III–RL–5 db/octave, III–RL–10 db/octave, and partial suppression filter system III–LP–1000–P.
FIG. 29. Rising loss system III–RL–10 db/octave without filter and with HPF–500, −1000.
FIG. 30. Resonant systems: III–RN–1100–3.9, −8.9.
FIG. 31. Resonant system III–RN–2000 and partial suppression filter system III–HP–2500–P.







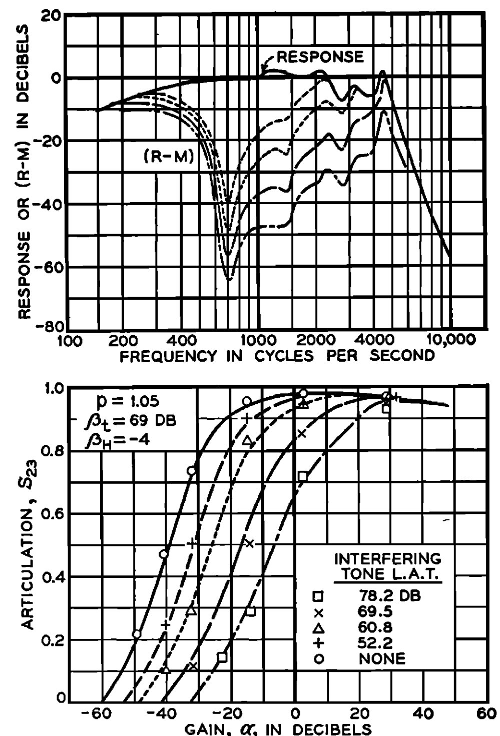


FIG. 32. Resonant system II–RN–1060, showing average speech sound articulation s_3 and also CVC syllable articulation S_3.
FIG. 33. System I with networks having different resonant frequencies but almost uniform damping.
FIG. 34. System I with networks having the same resonant frequency but different values of damping (db per millisecond).
FIG. 35. System I with 700 c.p.s. interfering tone at various levels above threshold. The upper plot shows R and also R–M where M is the masking caused by the pure tone.

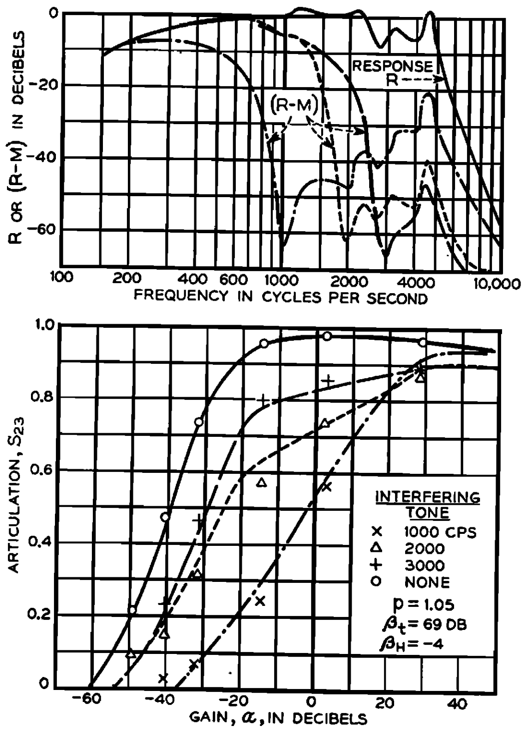


Fig. 36

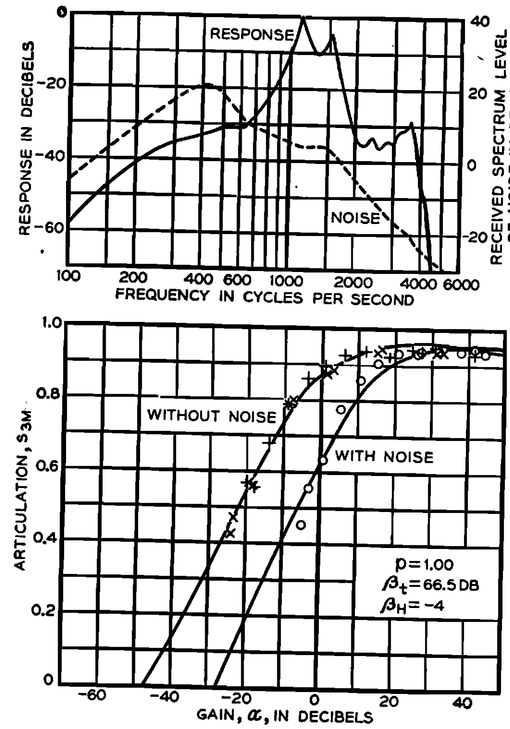


Fig. 38

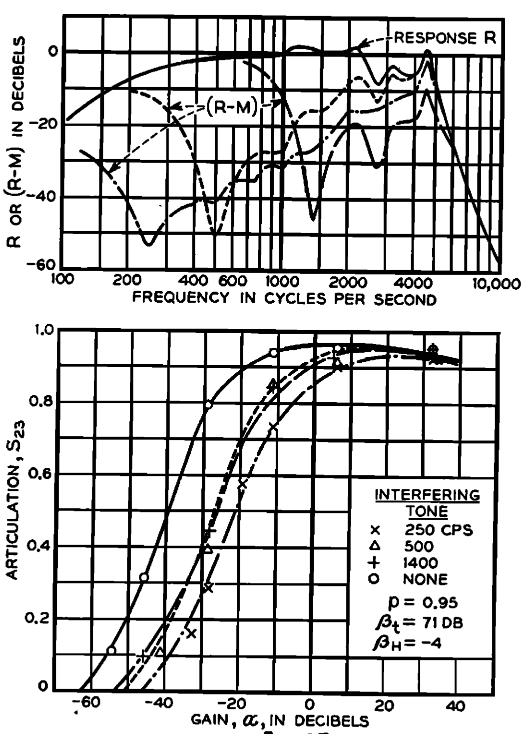


Fig. 37

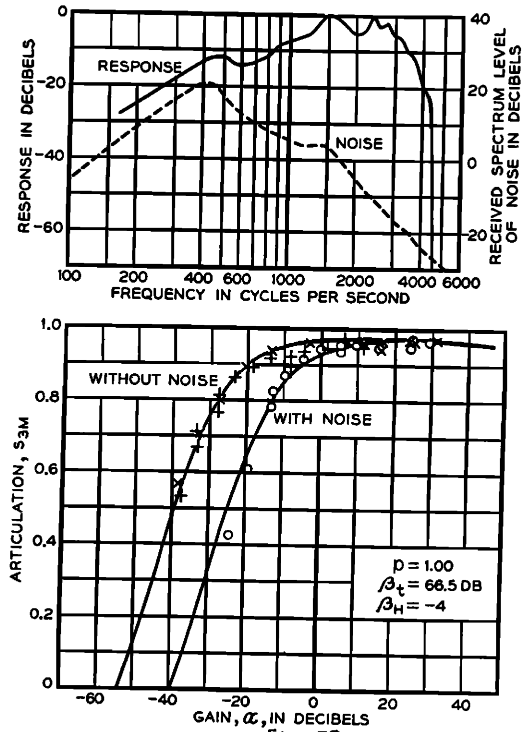


Fig. 39

FIG. 36. System I with interfering tone at 78.2 db above threshold, for three different tone frequencies.
FIG. 37. System I with interfering tone at 60.8 db above threshold, for three different tone frequencies.
FIG. 38. System III–HY–323 containing linear microphone (resembling No. 323 carbon microphone in certain properties), without and with a noise introduced electrically from room noise record.
FIG. 39. System III–HY–625 containing linear microphone (resembling No. 625 carbon microphone in certain properties), without and with a noise introduced electrically from room noise record.

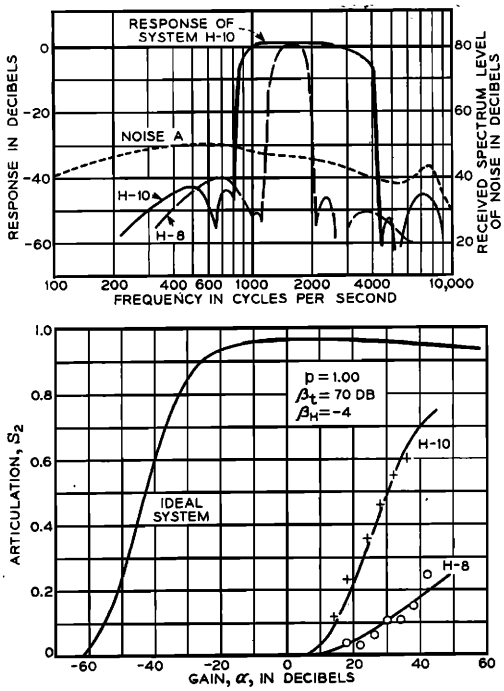


Fig. 40

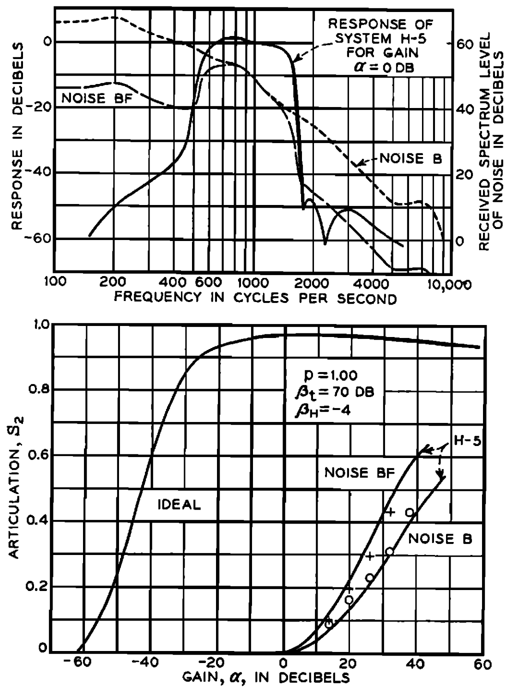


Fig. 42

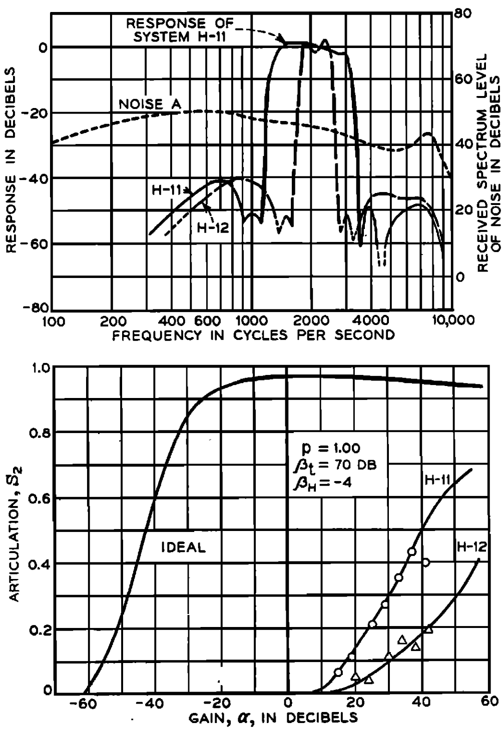


Fig. 41

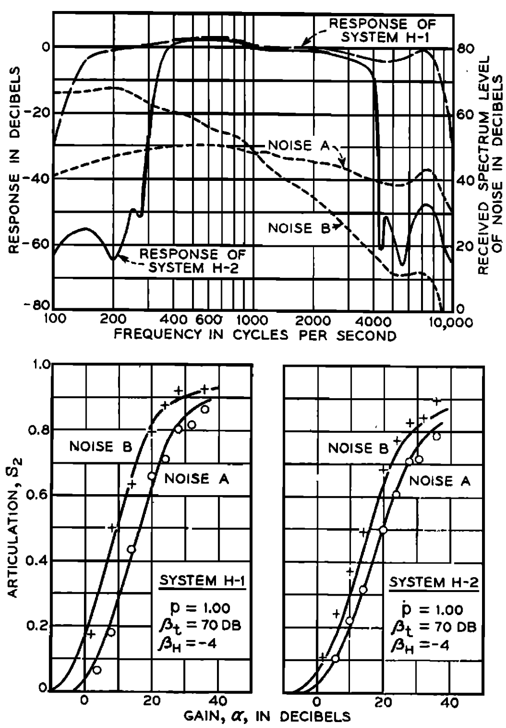


Fig 43

FIG. 40. Systems H–8, H–10 with noise *A*.
FIG. 41. Systems H–11, H–12 with noise *A*.
FIG. 42. System H–5 with noises *B* and *BF*.
FIG. 43. Systems H–1, H–2, with noises *A* and *B*.

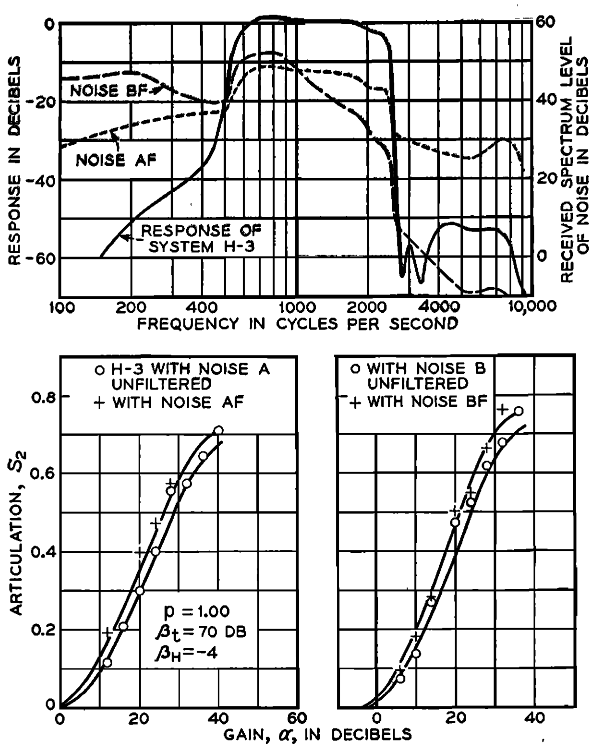


Fig. 44

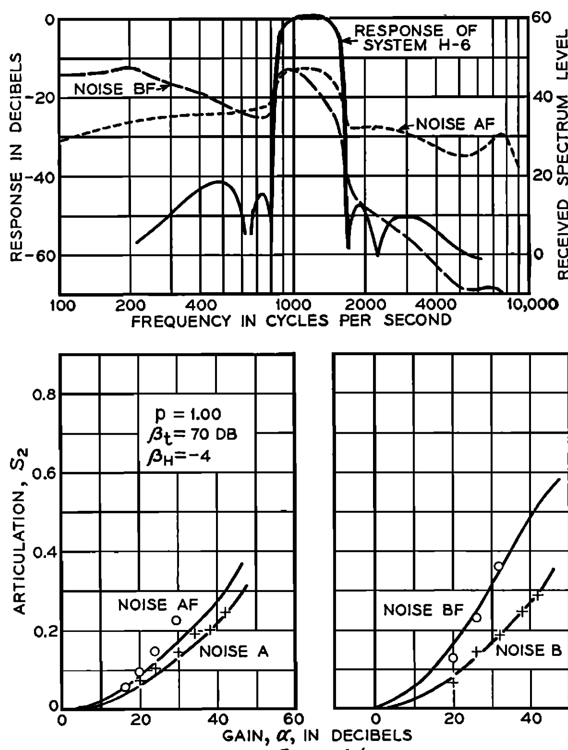


Fig. 46

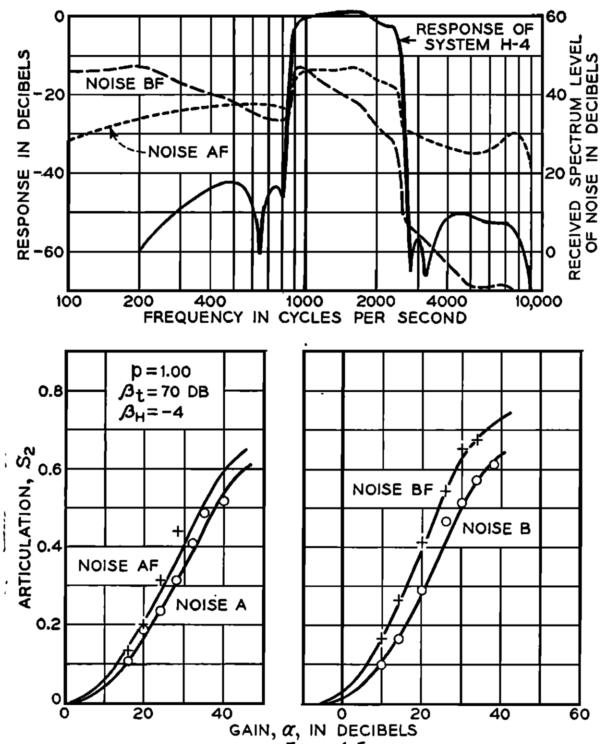


Fig. 45

FIG. 44. System H-3 with noises *A*, *AF*, *B*, *BF*. For noises *A* and *B*, see Fig. 43.

FIG. 45. System H-4 with noises *A*, *AF*, *B*, *BF*. For noises *A* and *B*, see Fig. 43.

FIG. 46. System H-6 with noises *A*, *AF*, *B*, *BF*. For noises *A* and *B*, see Fig. 43.

(hearing loss = 20 db) and softest caller ($\beta_t = 56$ db) the curve must be shifted 32 db to the right of the solid curve in Fig. 47, while for the best listener (hearing loss = −12 db) and the loudest talker ($\beta_t = 74$ db) the curve must be shifted 18 db to the left, as shown by the two dashed curves. For other talker-listener pairs the curve will lie between these extreme limits representing a range of 50 db. If we assume that a grade of transmission for which $A=\frac{1}{2}$ (which corresponds to $s_{3M}=0.88$, $S_3=0.68$, $S_{23}=0.71$, $S_2=0.79$ and $I=98$) is a tolerable one, then it is seen that this system will always give a better grade than this for this wide range of talker-listener pairs for $\alpha=0$—that is, with no additional attenuation or amplification. If one considered only the average talker-listener pair, an attenuation of 30 db could be introduced before the articulation s_{3M} drops to 0.88. The manner in which the talker and listener hold the transmitter and receiver modifies these general conclusions but in general when two persons fail to understand each other the speaker raises his voice and the listener holds the receiver more snugly. Both of these effects tend to push the right limiting curve to the left. One never has the ideal listening conditions considered in this case except in the laboratory because noise is always present.

Calculations will now be made for this system when room noise is present. D. F. Hoth[20] found that the relative spectrum level of room noise was the same for a very wide range of noise levels. D. F. Seacord[21] made measurements at a large number of telephone locations and found that the average room noise in residences corresponded to 43-db intensity level. The spectrum level corresponding to this average room noise is shown in Fig. 48 by the dashed line. The values of h, the attenuation of the noise going under the receiver cap to the ear, and the values of sidetone response R_s are given in this figure. The data were obtained by E. E. Mott and W. D. Goodale, Jr. of the Bell Telephone Laboratories. This, then, gives sufficient data to calculate A for any talker-listener pair for any intensity level of room noise since the relative spectrum level remains constant. At the bottom of Fig. 48 are shown such calculations for $\beta_t=68$ db and a listener of zero hearing loss, $\beta_H=0$. The intensity level of room noise as measured by a sound level meter is shown on each curve covering a range from 0 to 130-db intensity level.

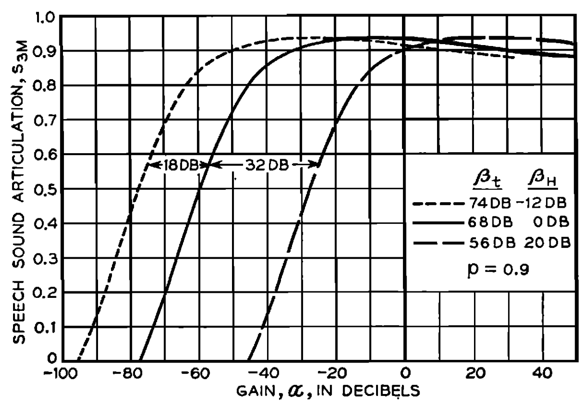


FIG. 47. Calculated articulation *vs.* gain for commercial system in quiet with range of talking level β_t and hearing loss β_H.

[20] D. F. Hoth, "Room noise spectra at subscribers' telephone locations," J. Acous. Soc. Am. **12**, 499 (1941).

[21] D. F. Seacord, "Room noise at subscribers' telephone locations," J. Acous. Soc. Am. **12**, 183 (1940).

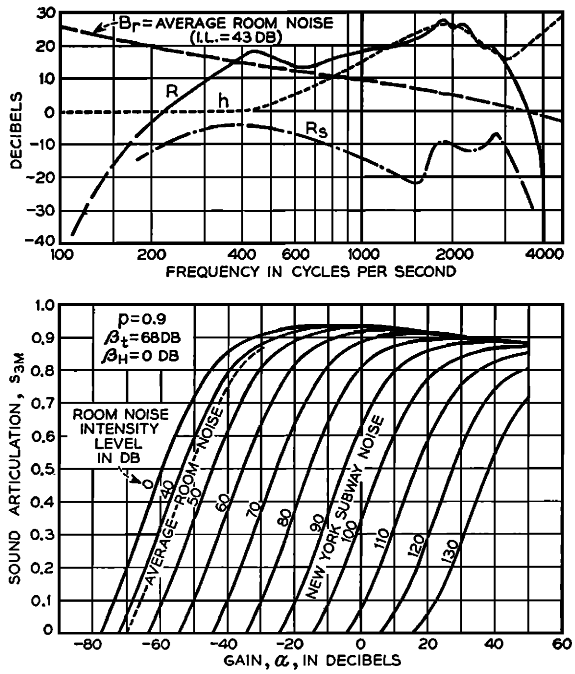


FIG. 48. Calculated articulation *vs.* gain for commercial system with room noise at listener's location. R = Response of system. B_r = Acoustic noise spectrum level. h = Attenuation of noise going under earphone cap to the ear. R_s = Sidetone response. $I.L.$ = Total intensity level.

The first curve marked 0 is the same as the solid curve in Fig. 47. It is seen that average room noise causes only about an 8-db shift from the curve for the no-noise case. All the listeners having a hearing loss of 8 db or less will have their threshold levels limited by the room noise. So when average room noise is present at the listener's end, then the effective acuity levels range from a hearing loss of 8 to a hearing loss of 20, or only 12 db instead of 32 db. But the talking levels cover the same range from 56 to 74 db. Therefore, when average room noise is present the range of curves is from one 2 db to the right of the one marked 0 to one which is 32 db to the right, or a range of 30 db. Then for a room noise of 55-db intensity level or greater, the threshold level of the listeners is determined entirely by the noise so variations above this are due only to the variations of the talking level or through a range of only 18 db. It is seen from the curves of Fig. 48 that the room noise can reach about 70-db intensity level before the transmission over this system reaches the tolerable limit of $s_{3M}=0.88$ for a 68-db talker and zero loss listener and when no amplification or attenuation is introduced into the system. It will be noticed that for noises of more than 100-db intensity level the tolerable limit of articulation chosen above is never reached even with amplifications as high as the ear can tolerate.

These calculations are for room noise at the listener's end of the line. If the noise is present at the talker's end, then the speech and noise are attenuated together as they go through the line to the receiver. The level difference at the receiving end is approximately the same as at the transmitting end. In a noise the talking level at a microphone close to the lips is approximately 100 db. As will be seen in the next calculation, unless the room noise level is greater than 60 db (that is, less than 40 db below the talking level) no reduction in articulation will occur. For greater levels of room noise at the transmitting end, one must find the spectrum level of the transmitted noise when it arrives at the listener's ear and then combine this with the noise at the receiving end to find the resultant noise spectrum.

In the next calculation is considered the effect of a noise having spectrum level constant with frequency upon the transmission from an ideal filter system passing frequencies from 125 to 5700 c.p.s. The noise is also limited to this frequency range. Thus one can compare the transmission for various values of the average speech intensity level β_s, and of the noise intensity level β_N, at the listener's ear. The results of the calculation are shown in Fig. 49. The ordinates of each solid line give syllable articulation S_3 for one value of β_N and the abscissas give the received speech intensity level β_s. Each dashed curve represents a condition in which β_s differs from β_N by a constant number of db, namely, -10, 0, 10, 20, 30, and 40 db. Thus each dashed curve shows the effect of increasing and decreasing the gain in a receiving set which amplifies or attenuates both the noise and the speech. These curves show that under the assumed conditions increasing the gain so that the speech at the listener's ear is more than about 70 db above the threshold is harmful in all cases.

A third application will show how to calculate the shift of the articulation *versus* gain curve due to changing some of the elements in the system. The gain curve shift as used in this paper will now be defined. Consider a system which has a response R *versus* f. Now introduce a network into this system that changes the response to R' *versus* f and other distortions that change H to H'. The articulation index A of the unchanged system will be designated

$$A = H\cdot E\cdot F\cdot V(x_V)$$

and that of the changed system,

$$A' = H'\cdot E'\cdot F'\cdot V(x'_V).$$

If a curve of articulation index *versus* gain is calculated for the two systems, then if $\alpha_{\frac{1}{2}}$ is the gain to reach $A=\frac{1}{2}$ for the first system and $\alpha_{\frac{1}{2}}'$ the gain to reach $A'=\frac{1}{2}$ for the second system, then $\alpha_{\frac{1}{2}}'-\alpha_{\frac{1}{2}}$ is the gain curve shift.

For the unchanged system

$$\alpha = x_V+\alpha_0+\gamma\phi(\bar{R}_1-\bar{R}_4).$$

Let the primed quantities represent the values for the changed system or

$$\alpha' = x'_V+\alpha_0'+\gamma'\phi'(\bar{R}_1'-\bar{R}_4').$$

Then $\alpha'-\alpha$ represents the shift of one articulation index *versus* gain curve from the other and its value

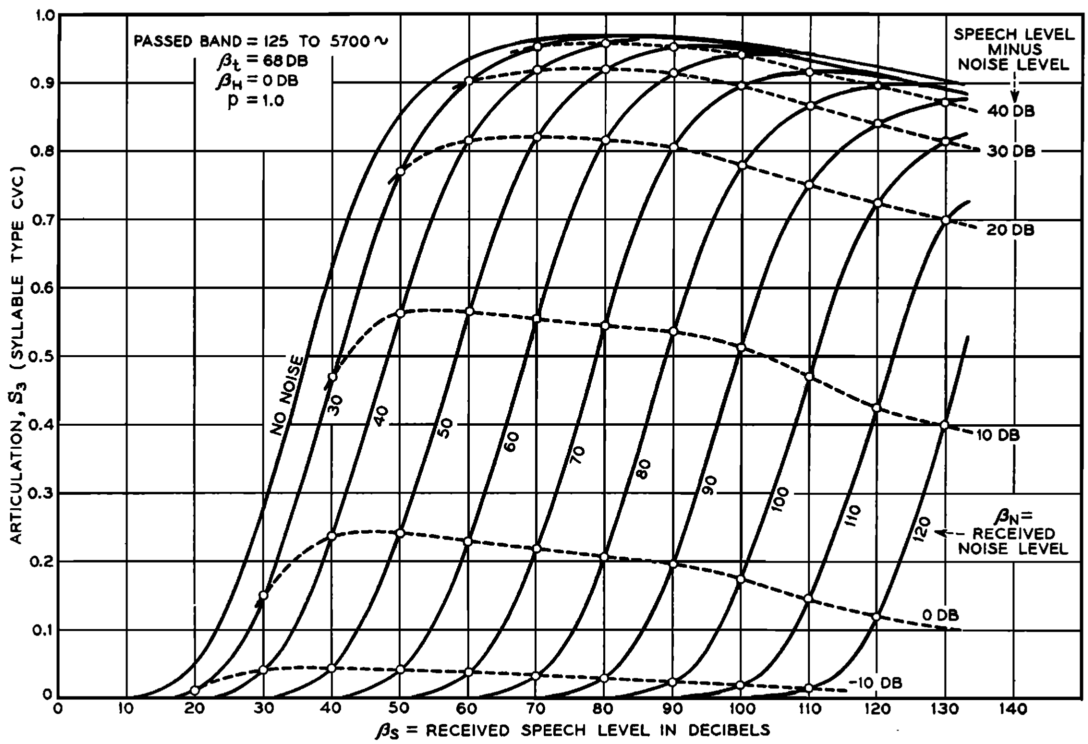


FIG. 49. Calculated articulation *vs.* gain for an ideal filter system with a noise having uniform spectrum level at all passed frequencies.

TABLE XXVII. Calculation of gain curve shift for singly resonant system.

f_0	F	$\bar{R}_1$	$1/2F$	x_V	ϕ	$\gamma\phi(\bar{R}_1-\bar{R}_4)$	Gain curve shift in db
250	0.909	−31.7	0.550	30.3	0.97	1.1	34.7
500	0.908	−16.2	0.551	30.4	0.97	9.8	28.0
1000	0.911	−10.7	0.548	30.2	0.97	10.1	22.6
2000	0.926	−10.3	0.540	30	0.97	6.8	18.7
4000	0.933	−16.6	0.536	29.8	0.97	3.1	21.1

TABLE XXVIII. Calculation of gain curve shift for doubly resonant system.

f_0	F	$\bar{R}_1$	$1/2F$	x_V	ϕ	$\gamma\phi(\bar{R}_1-\bar{R}_4)$	Gain curve shift in db
250	0.795	−57.9	0.63	32.8	0.98	1.9	64.2
500	0.590	−19.2	0.847	46.6	1	30.3	67.7
1000	0.599	−14.2	0.835	45.5	1	26.4	57.6
2000	0.655	−13.8	0.752	40.2	1	21.8	47.2
4000	0.678	−20.2	0.737	39.0	1	20.6	51.4

at the ordinate $A=\frac{1}{2}$ is the value of the gain curve shift or $(\alpha'-\alpha)_{\frac{1}{2}}$, which is given by

$$(\alpha'-\alpha)_{\frac{1}{2}}=(x'_V-x_V)_{\frac{1}{2}}+(\bar{R}_1-\bar{R}_1')+[\gamma'\phi(\bar{R}_1'-\bar{R}_4')-\gamma\phi(\bar{R}_1-\bar{R}_4)] \quad (78)$$

since $\alpha_0'-\alpha_0=\bar{R}_1-\bar{R}_1'$.

To illustrate, consider the gain curve shift due to introducing singly resonant elements into an ideal system (flat response $R=0$). The response R' is given by

$$R'=10\log\{1+745(f_0/\Delta)^2[(f/f_0)-(f_0/f)]^2\}, \quad (79)$$

where the resonant frequency f_0 is expressed in kilocycles and the damping constant Δ is expressed in db per millisecond. Consider systems where $f_0/\Delta=0.4$. For such resonant systems $R=0$ at $f=f_0$. For the ideal system $\bar{R}_1=\bar{R}_4=R=0$ at all frequencies. For this ideal system the effective gain x_V for $A=\frac{1}{2}$ is 28.4. Therefore, the gain curve shift $(\alpha'-\alpha)_{\frac{1}{2}}$ is given by

$$(\alpha'-\alpha)_{\frac{1}{2}}=x'_V-28.4-\bar{R}_1'+\gamma'\phi'(\bar{R}_1'-\bar{R}_4'). \quad (80)$$

For the resonant system x'_V is determined by the relation

$$F'\cdot V(x'_V)=\tfrac{1}{2} \quad \text{or} \quad V(x'_V)=1/2F. \quad (81)$$

The values F', $\bar{R}_1'$, $\bar{R}_4'$, γ' and ϕ' were calculated from such response curves and are given in Table XXVII (omitting the primes used in the text). The values of gain curve shift given in the last column were calculated from Eq. (80).

It is seen that the value of F is approximately the same for all five resonant systems, from which it follows that the maximum value of articulation for each of these systems is about the same. Thus the resonant frequency can be placed anywhere between 250 to 4000 c.p.s. without reducing the factor F below 0.90. However, the values of gain curve shift are quite different, varying from 18.7 to 34.7 db. If the diaphragm of the transmitter or receiver is considered as the resonant element, then the gain curve shift is equal in magnitude to the gain that would be achieved if the instrument had the same efficiency at all frequencies as for that corresponding to the resonant frequency.

Similar calculations are shown in Table XXVIII for the case of two independent resonant elements in the system, each having the same resonant frequency. In other words, the response R in db is double that for the singly resonant system. The resonant elements may be considered as diaphragms in the microphone and the receiver.

If one resonant peak is placed at 1000 c.p.s. and the other at 2000 c.p.s., the value of $F=0.84$ and the gain curve shift is only 14.7 db.

If four resonant elements of this type are arranged so that the resonant peaks are at 500, 1000, 2000, and 4000 c.p.s., then $F=0.978$ and the gain curve shift is only 4.7 db. In this case one has essentially a high quality system. In the modern design of telephone microphones and receivers the diaphragms are made essentially multiresonant so that they pass efficiently a wide band of frequencies.

A fourth application of the present method will show how to calculate articulation for deafened persons. This application will be described in a subsequent paper with a comparison between calculations and observations.

23. ACKNOWLEDGMENTS

In the text of the present paper, and in the references, the names of various persons have been indicated who have taken an active part in planning, performing, and interpreting the articulation tests. It may be found that some names have been omitted which should equally have been included. The authors desire to acknowledge the achievements of all of the individuals who have contributed to the material used in this study.

The authors wish particularly to thank Miss Priscilla A. Pecon for her able work in the many computations here reported, and to thank Miss Jane A. Otto for invaluable assistance in editing the paper.

APPENDIX 1: THE LOUDNESS FUNCTIONS G_1 AND G_4

In this appendix the two loudness functions G_1 and G_4 used in the calculation of the articulation are derived. The equations defining G_1 and G_4 are, from Eqs. (23) and (26) in the text,

$$10^{\bar{R}_1/10}=\int_0^\infty G_1 10^{R/10}df \quad (23)$$

and

$$10^{\bar{R}_4/40}=\int_0^\infty G_4 10^{R/40}df. \quad (26)$$

For an ideal system, R is independent of frequency and equal to $\bar{R}_1$ or $\bar{R}_4$ so

$$\int_0^\infty G_1 df = \int_0^\infty G_4 df = 1. \qquad (28)$$

The value of G_1 will be determined first. Consider an ideal high quality system, having the same response R at every frequency. Consider also several ideal filter systems with various cut-off frequencies, having the same response R at every transmitted frequency and the response $-\infty$ at all other frequencies. The gain of each of these systems, having these values of response, is designated as gain $\alpha=0$ db. Any other gain may be specified by assigning a value to α.

Heretofore in this paper the symbol α_0 has been used to designate for any system the value of gain at which the received speech is at the threshold of audibility. In this appendix, however, a_0 will be used to designate the threshold adjustment for the unfiltered system only The gain corresponding to received speech at threshold for any of the filtered systems will be designated as $\alpha_0+\Delta\alpha$. Thus let the ideal unfiltered system have the gain α_0, so that speech is at threshold. Then if an ideal filter is inserted, the gain must be increased by the amount $\Delta\alpha$ in order that the filtered speech may be at threshold.

Then for low pass filter systems

$$10^{(\bar{R}_1+\alpha_0)/10} = \int_0^{f_c} 10^{(R+\alpha_0+\Delta\alpha)/10} G_1 df.$$

But in the passed region $\bar{R}_1 = R$ so

$$10^{-\Delta\alpha/10} = \int_0^{f_c} G_1 df. \qquad (24)$$

Similarly for high pass filter systems

$$10^{-\Delta\alpha/10} = \int_{f_c}^\infty G_1 df = 1 - \int_0^{f_c} G_1 df. \qquad (25)$$

These values $\Delta\alpha$ are the number of decibels the filter system must be raised from α_0 to make the received speech audible.

Experimental values of $\Delta\alpha$ were obtained from threshold observations upon twenty-six filter systems which had been used in articulation tests. Eighteen of these systems were in the 1928–1929 group, ten of these having the responses shown in Figs. 25 and 27. The remaining eight systems were in the 1935–1936 group and had the responses shown in Figs. 24 and 26. It will be seen that these were not ideal filter systems, hence corrections were necessary in order to convert the observed attenuation for threshold for each actual filter into the gain change $\Delta\alpha$ corresponding to the cut-off frequency f_c of an ideal filter.

For each actual filter system, and for the corresponding actual high quality (i.e., unfiltered) system, the average response in the transmitted region was found for the gain adjustment at which the received speech was at the threshold of audibility. The difference between these two average responses was taken to be the desired gain change $\Delta\alpha$. It was necessary to adopt some form of frequency weighting in finding the average response, which was done in the manner about to be described. It was necessary also to select some value of frequency to be regarded as the cut-off frequency for any particular filter. The frequencies so selected were the same as those adopted in the study of maximum values of articulation observed for the filters, as recorded in Tables II and IV. Although somewhat different frequencies might more properly have been chosen for the threshold data, the effects of such differences are regarded as small in comparison with the uncertainties involved in threshold observations.

If the function G_1 were known, the average response R_{AV} for the entire passed frequency region would be given by the equation

$$10^{R_{AV}/10} = \frac{\int_0^{f_c} G_1 10^{R/10} df}{\int_0^{f_c} G_1 df}. \qquad (82)$$

Thus the frequency weighting which should be employed in determining R_{AV} depends upon the same function G_1 which we are seeking. Here we have recourse to successive approximations.

As a first approximation, the average response R_{AV} was estimated by inspecting the response *versus* frequency characteristic of each system. Using these tentative values of R_{AV}, preliminary values of G_1 were obtained which were then used to determine R_{AV} more accurately. This procedure can be repeated as often as may be needed. Such values of R_{AV} are given in column 3 of Table XXIX. Each of these values corresponds to an over-all gain adjustment which has been designated as $\alpha=0$ db in the figure showing the response R *versus* f.

For each system, column 4 of Table XXIX gives the experimental value of the attenuation required to bring the received speech to the threshold level. Subtracting the attenuation in column 4 from R_{AV} in column 3 we obtain in column 5 the value R_{AV}', which is the average response when speech is at threshold. If we let R_0 be the value of R_{AV}' for the unfiltered system, then for any filter system $R_{AV}'=R_0+\Delta\alpha$.

In column 6 of Table XXIX the value of $\Delta\alpha$ is the difference between two values in column 5, namely between $R_{AV}'=R_0+\Delta\alpha$ for each filter system and $R_{AV}'=R_0$ for the unfiltered system. Finally, the values of $\Delta\alpha$ are substituted in Eqs. (24) and (25) to obtain column 7, which gives the values of $\int_0^{f_c} G_1 df$.

The determination of $\Delta\alpha$ in the manner just described is based upon the assumption that the same talking level β_t was used in the threshold test of a filtered system as in the companion test of the unfiltered system. All of the 1928–1929 (i.e., System II) data have been entered in Table XXIX on this basis, with $\beta_t=69$ db. The 1935–1936 (i.e., System III) low pass filter data in Table XXIX column 4 have been referred to the talking level $\beta_t=69$ db as in Fig. 24, whereas the corresponding high pass filter data have been referred to $\beta_t=68.5$ db as in Fig. 26.

The values of $\int_0^{f_c} G_1 df$ *versus* f_c from Table XXIX are plotted as discrete points in Fig. 50. The solid curve drawn through the array of points shows the function $\int_0^{f_c} G_1 df$ used in this paper, the slope of which is the function G_1 also shown in Fig. 50 and given in Table XXX column 3. The same pair of functions are shown in Fig. 9. As a check these functions were used with Eqs. (24) and (25) to calculate values of $\Delta\alpha$, which are shown in Fig. 51 by the pair of curves together with the observed values of $\Delta\alpha$ from Table XXIX plotted as points. Although the curves fit the points only fairly well, the fit was accepted as adequate in view of the success attained by the functions of this paper (including the G_1 function) in the calculation of articulation.

TABLE XXIX. Derivation of the function $\int_0^{f_c} G_1 df$ from threshold observations upon filtered speech.*

(1) System designation	(2) f_c c.p.s.	(3) R_{AV} for $\alpha=0$ db	(4) A_0 db to threshold	(5) R_{AV}' for threshold db	(6) $\Delta\alpha$ db	(7) $\int_0^{f_c} G_1 df$
II	6750	−2.8	58.2	−61.0	0.0	1.000
II–LP–7000	6750	−2.8	58.2	−61.0	0.0	1.000
II–LP–5500	5440	−2.8	58.2	−61.0	0.0	1.000
II–LP–4500	4300	−2.8	58.2	−61.0	0.0	1.000
II–LP–3750	3625	−2.8	57.9	−60.7	0.3	0.933
II–LP–3250	3185	−2.7	58.2	−60.9	0.1	0.977
II–LP–2850	2800	−2.7	57.8	−60.5	0.5	0.891
II–LP–2450	2400	−3.1	57.7	−60.8	0.2	0.955
II–LP–1950	1950	−3.1	58.2	−61.3	−0.3	1.072
II–LP–1500	1460	−3.3	55.0	−58.3	2.7	0.537
II–LP–1000	990	−4.0	52.0	−56.0	5.0	0.316
II–LP–750	755	−4.0	48.9	−52.9	8.1	0.155
II	260	−2.8	58.2	−61.0	0.0	0.000
II–HP–250	320	−2.8	58.2	−61.0	0.0	0.000
II–HP–500	530	−3.0	58.0	−61.0	0.0	0.000
II–HP–750	810	−3.3	56.2	−59.5	1.5	0.292
II–HP–1000	1030	−3.8	54.8	−58.6	2.4	0.425
II–HP–1500	1525	−5.6	51.3	−56.9	4.1	0.611
II–HP–1900	1915	−5.7	49.2	−54.9	6.1	0.754
II–HP–2900	2865	−8.2	44.2	−52.4	8.6	0.862
III	6500	−3.8	56.3	−60.1	0.0	1.000
III–LP–4500	4400	−4.0	55.1	−59.1	1.0	0.794
III–LP–2850	2885	−3.5	55.6	−59.1	1.0	0.794
III–LP–1850	1905	−3.1	55.2	−58.3	1.8	0.661
III–LP–1850	1905	−3.1	53.3	−56.4	3.7	0.427
III–LP–1000	1040	−3.5	46.2	−49.7	10.4	0.091
For the above tests $\beta_t=69.0$ db; below, $\beta_t=68.5$ db.						
III	285	−3.8	55.8	−59.6	0.0	0.000
III–HP–500	565	−4.7	52.6	−57.3	2.3	0.411
III–HP–1000	1025	−5.4	52.4	−57.8	1.8	0.339
III–HP–1500	1425	−9.9	46.1	−56.0	3.6	0.563
III–HP–1500	1425	−9.9	44.2	−54.1	5.5	0.718
III–HP–2500	2520	−8.2	40.8	−49.0	10.6	0.913

* The entries in columns 6 and 7 for II–LP–1950 are interpreted in Figs. 50 and 51 as though $\Delta\alpha=0.0$ db.

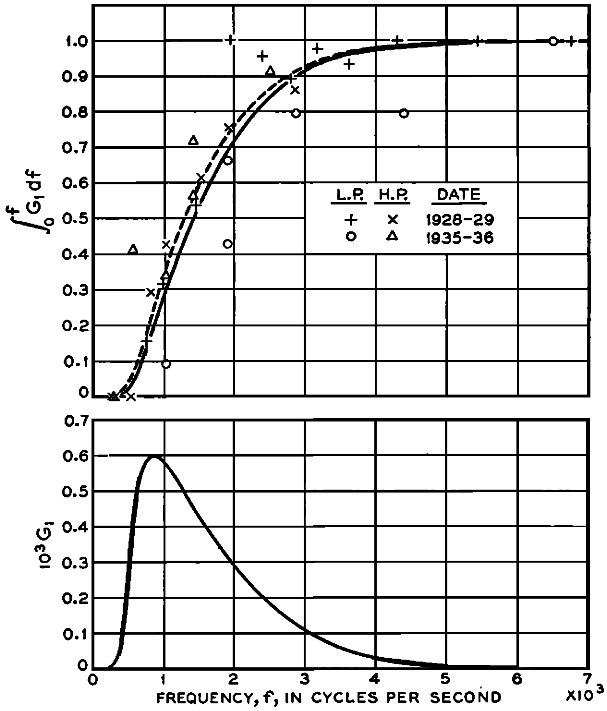


FIG. 50. Derivation of $\int_0^f G_1 df$ and G_1 from filtered speech thresholds. (See also Fig. 9 and Table VIII.) The discrete points in the upper plot show values of $\int_0^{f_c} G_1 df$ *vs.* f_c derived in Table XXIX from observed thresholds.

Except for the time involved in continuing the revision of the functions, it would have been better as the next approximation to represent the observed points by a revised graph of $\int_0^{f} G_1 df$ *versus* f about as that shown by the broken line in Fig. 50. The slope of this revised integral curve at each frequency would then constitute a revised G_1 function which would more accurately represent the observations of speech thresholds than does the G_1 function used in this paper.

In the 1928–1929 threshold tests the procedure was to find a gain adjustment such that at this gain the speech was heard by the listener, but that with the attenuation increased by 5 db the speech was inaudible. The gain halfway between these two points was taken as the threshold adjustment. The average of these adjustments was found for the entire articulation test crew, so that all voices both male and female and all ears entered into the average, although it is not certain whether or not every voice-ear combination was used. For each system, filtered or unfiltered, the result was expressed as an integral number of db of attenuation required to bring the received speech to threshold. This adoption of integral numbers implies a smoothing operation to the extent of a fraction of one db, which results in a less scattered set of points in Fig. 50 representing the earlier observations as compared with the later observations which received no smoothing.

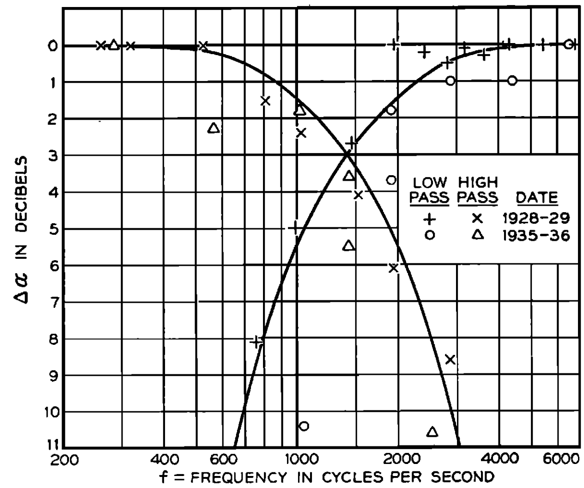


FIG. 51. Gain change $\Delta\alpha$ required for speech to remain at threshold when cut-off frequency is changed from 0 or ∞ to f. Points observed, curves calculated. Values of $\Delta\alpha$ are zero or positive but are plotted increasing downward.

In the 1935–1936 tests the procedure was to change the attenuation by steps of 2 db in determining the threshold. The attenuation A_0 was found at which speech was audible in about three-fourths of the trials, but such that with (A_0+2) db the speech was heard in much less than one-half of the trials. The attenuation A_0 was adopted as the adjustment for threshold. Care was taken to obtain this adjustment for each talker-listener combination of the articulation test crew. For any given filter, also for the unfiltered system, the average was found of the values of A_0 obtained by a listener for the various voices. Thus there was a group of seven or eight (for one crew, six) listener's averages of which the mean or grand average was adopted in Table XXIX as the observed attenuation A_0 for threshold for the entire crew. The r.m.s. deviation of one listener's average from the grand average for the crew was typically about 4 db.

The accuracy of repetition of the 1935–1936 threshold tests can be judged from three different systems. As column 4 of Table XXIX indicates, in two instances a repetition of the test upon a filter system resulted in a second grand average A_0 which differed from the first value by 1.9 db. In the tests of the unfiltered system, three values of the grand average A_0 showed a total spread of 2.6 db; only the mean of these results is given in Table XXIX. Each of these repetitions involves a change of only one or two crew members. Where it was possible, each value of $\Delta\alpha$ comes from a pair of tests (of the filter system, and of the unfiltered system) made by the same crew.

The function G_1 can be calculated from loudness relations as follows.

In a paper by Fletcher and Munson[22] it was shown that the loudness N of thermal noise could be calculated from the formula

$$N=\int_0^{100} Q(Z)dx, \tag{83}$$

where Z, Q, and x are functions of the frequency, which will now be defined.

The value x is the percent of the total nerve endings in the inner ear which has been passed over by the maximum stimulation as the impressed frequency of a tone goes from zero to f. This relationship between x and f has been determined from measurements of frequency discrimination data, from the Stevens and Volkmann scale of pitch in mels, from critical band-width data and from direct anatomical measurements. As stated in the paper, the quantity Z is defined in terms of the sum of three other quantities, namely

$$Z=B+\kappa-\beta_0. \tag{37}$$

All of these quantities vary as the frequency varies. The quantity B is the spectrum intensity level at each frequency.

It has been found for systems having an approximately flat response such as systems I, II, or III that the loudness of speech at the receiving end of the system varies with the db above threshold $\alpha-\alpha_0$ as shown in Fig. 52. The points × give unpublished data obtained by W. A. Munson and the points □ give data from the book, *Speech and Hearing*, (see reference 9, p. 232, Fig. 111). The curve was taken as the best fit of the

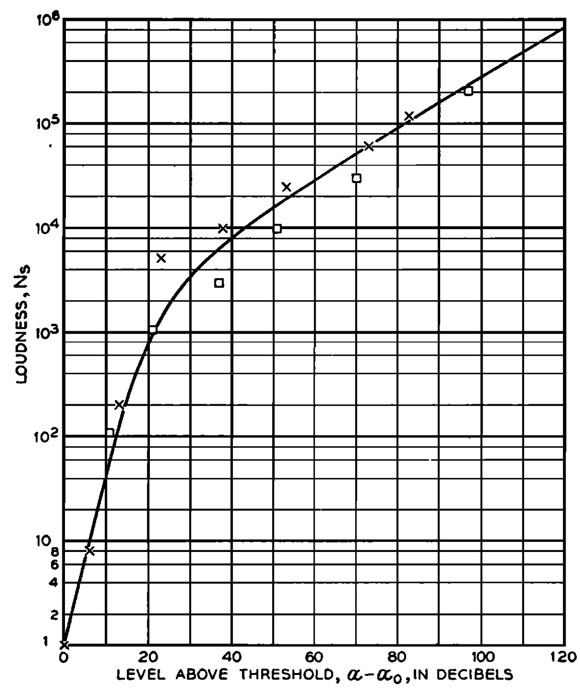


FIG. 52. Loudness of undistorted speech *vs.* level above threshold. Points observed, curve drawn to represent both sets of data.

[22] H. Fletcher and W. A. Munson, "Relation between loudness and masking," J. Acous. Soc. Am. 9, 1 (1937).

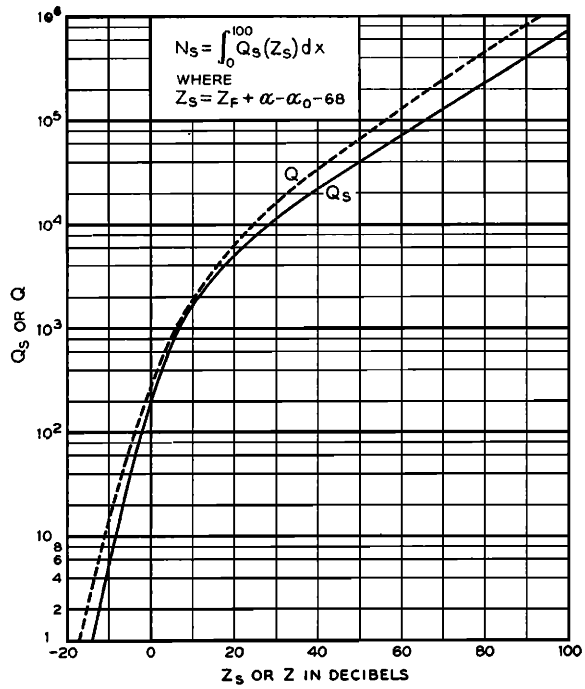


FIG. 53. The functions Q_s *vs.* Z_s and Q *vs.* Z.

combined sets of data. From this curve values of Q_s were determined such that for speech

$$N_s=\int_0^{100} Q_s(Z_s)dx, \tag{84}$$

where for a flat response system

$$Z_s=Z_F+\alpha-\alpha_0-68. \tag{85}$$

Here, as explained in Section 10, Z_F is the level above threshold of a critical band of speech when the speech is received at the optimum level for interpretation. The values of Q_s satisfying the loudness data for speech are shown in Fig. 53. When used in Eq. (84) the observed data shown in Fig. 52 will be calculated correctly. For comparison, Fig. 53 shows also the values of Q for thermal noise as given in the Fletcher and Munson paper.[22] It will be seen that at levels of Z_s near zero the value of Q_s can be represented approximately by

$$Q_s(Z_s)=K_1 10^{Z_s/10} \tag{86}$$

and at levels above $Z_s=35$, Q_s can be represented by

$$Q_s(Z_s)=K_4 10^{Z_s/40}. \tag{87}$$

If these values of Q_s and the value of Z_s from Eq. (85) are substituted in Eq. (84) the required formula for calculating the loudness of speech from a transmission system is obtained. However, the loudness N_s is given in terms of the variable x instead of f.

It is known that κ is related to the critical band width $(\Delta f)_c$ by the equation

$$(\Delta f)_c=10^{\kappa/10}. \tag{88}$$

Also, if $(\Delta x)_c$ is the critical band of x—that is, the critical width of the nerve patch in the inner ear corresponding to the critical frequency band width $(\Delta f)_c$, then the

TABLE XXX. Values of the G functions and of B_s, β_0, β_0', $\kappa-\beta_0'$, and Z_F.

(1) f	(2) $\int_0^f G_1 df$	(3) $G_1\times10^3$	(4) $G_4\times10^3$	(5) $\int_0^f G_4 df$	(6) Average speech B_s	(7) Peaks ΔB_s	(8) $B_s+\Delta B_s$ $-10\log G_1$ -70	(9) Pure tone β_0	(10) For listening to speech β_0'	(11) Critical bands of speech in db κ	(12) $\kappa-\beta_0'$	(13) $B_s+\Delta B_s$ $+8+\kappa-\beta'_0$ Z_F
100	0.0000	0.00001	0.019	0.000	34.7	7.5		41.4	(48)	19.4	−28.6	21.6
200	0.0000	0.001	0.090	0.009	38.3	7.8		26.9	(34)	17.2	−16.8	37.3
300	0.0009	0.016	0.186	0.019	39.5	8.6		18.6	(26)	17.0	−9.0	48.1
400	0.0056	0.078	0.266	0.041	39.8	9.3	20.2	13.4	20.2	17.0	−3.2	53.9
500	0.0221	0.252	0.371	0.075	39.4	9.8	15.2	9.9	15.2	17.0	1.8	59.0
600	0.057	0.448	0.414	0.112	38.5	10.2	12.2	7.4	12.2	17.2	5.0	61.7
700	0.107	0.557	0.430	0.153	36.8	10.6	9.9	5.7	9.9	17.3	7.4	62.8
800	0.165	0.593	0.414	0.195	34.9	10.9	8.1	4.6	8.1	17.6	9.5	63.3
900	0.224	0.599	0.401	0.233	33.3	11.3	6.8	4.2	6.8	17.8	11.0	63.6
1000	0.284	0.587	0.386	0.279	31.8	11.5	5.6	4.0	5.6	18.0	12.4	63.7
1250	0.421	0.512	0.345	0.372	28.7	12.1	3.7	3.7	3.7	18.5	14.8	63.6
1500	0.539	0.436	0.301	0.453	26.1	12.7	2.4	2.7	2.4	19.0	16.6	63.4
1750	0.639	0.358	0.263	0.522	24.0	13.2	1.7	1.6	1.7	19.5	17.8	63.0
2000	0.720	0.290	0.233	0.582	22.1	13.7	1.2	0.1	0.6	19.9	19.3	63.1
2500	0.839	0.187	0.179	0.685	19.0	14.6	0.9	−2.1	−0.6	20.8	21.4	63.0
3000	0.914	0.113	0.140	0.764	16.4	15.1	1.0	−3.4	−1.2	21.5	22.7	62.2
3500	0.957	0.060	0.104	0.823	14.2	15.4	1.8	−3.7	−1.0	22.3	23.3	60.9
4000	0.976	0.030	0.077	0.870	12.3	15.5	3.0	−3.2	−0.1	23.1	23.0	58.9
4500	0.987	0.016	0.059	0.904	10.6	15.5	4.1	−2.0	1.0	23.7	22.7	56.3
5000	0.993	0.008	0.044	0.931	9.1	15.5	5.6	0.1	2.8	24.4	21.6	53.7
6000	0.998	0.0014	0.023	0.963	6.5			7.1	7.1	25.7	18.6	47.1
7000	0.999	0.0003	0.012	0.980	4.3			12.1	12.1	26.8	14.7	42.0
8000	1.000	0.0001	0.008	0.988	2.4			15.1	15.1	27.7	12.6	38.0
9000	1.000	0.0000	0.006	0.994	0.7			17.5	17.5	28.5	11.0	34.3
10000	1.000	0.0000	0.004	1.000	−0.7			19.5	19.5	29.2	9.7	32.0

slope

$$\frac{\Delta x}{\Delta f}=\frac{(\Delta x)_c}{(\Delta f)_c}=(\Delta x)_c 10^{-\kappa/10}. \quad (89)$$

From the theory of hearing $(\Delta x)_c$ is interpreted as a constant for all frequency regions and has been found to be equal to 1.56 percent, which is equivalent to about $\frac{1}{2}$ mm along the basilar membrane. So the values of the slope $\Delta x/\Delta f$ can be obtained from Eq. (89) and the values of x can be obtained from the equation

$$x=1.56\int_0^f 10^{-\kappa/10}df. \quad (90)$$

From Eqs. (84) and (86), the equation for calculating loudness for the low levels of received speech then becomes

$$N_s=1.56K_1\int_0^\infty 10^{(Z_s-\kappa)/10}df.$$

For a flat response system it was seen that Z_s is given by Eq. (85). In general, for a system having a non-uniform response R the value of Z_s is given by

$$Z_s=Z_0+R+\alpha+(\beta_t-68)$$

where Z_0 is the db above threshold when α and R are zero and $\beta_t=68$. The value of Z_0 is given by

$$Z_0=B_s+\Delta B_s+\kappa-\beta_0'$$

where the values of B_s, ΔB_s, κ and β_0' are those given in Table XXX.

Then

$$N_s=1.56K_1\int_0^\infty 10^{(Z_0+R+\alpha+\beta_t-68-\kappa)/10}df. \quad (91)$$

Consider two telephone systems 1 and 2. System 1 has a response R which is different at each frequency but known from the response curve for the system. System 2 is an ideal flat response system with $R=0$ at all frequencies. Let the amplification α_2 in system 2 be set so that the speech received from this system sounds equally loud to that received from system 1 with $\alpha=0$. Under these conditions the loudness values N_s from Eq. (91) are equal or

$$\int_0^\infty 10^{(Z_0+\alpha_2-\kappa)/10}df=\int_0^\infty 10^{(Z_0+R-\kappa)/10}df. \quad (92)$$

Then the value of α_2 is the desired value of $\bar{R}_1$ so that

$$10^{\bar{R}_1/10}=\frac{\int_0^\infty 10^{(Z_0-\kappa)/10}\cdot 10^{R/10}df}{\int_0^\infty 10^{(Z_0-\kappa)/10}df}. \quad (93)$$

Comparing Eqs. (93) and (23) it is seen that the value of G_1 is given by

$$G_1=\frac{10^{(Z_0-\kappa)/10}}{\int_0^\infty 10^{(Z_0-\kappa)/10}df}. \quad (94)$$

In a similar way at the high levels the condition for equality of loudness is

$$\int_0^\infty 10^{[(Z_0+\alpha_2)/40]-[\kappa/10]}df=\int_0^\infty 10^{[(Z_0+R)/40]-[\kappa/10]}df \quad (95)$$

and in this case α_2 becomes $\bar{R}_4$ so that

$$G_4=\frac{10^{(Z_0-4\kappa)/40}}{\int_0^\infty 10^{(Z_0-4\kappa)/40}df}. \quad (96)$$

Comparing Eqs. (94) and (96) it is seen that

$$G_4=G_1^{\frac{1}{4}}10^{-3\kappa/40} \quad \text{(constant)}. \quad (97)$$

The constant can be determined by the requirement that

$$\int_0^\infty G_4df=1.$$

Thus it is seen that the values of G_4 can be derived from the values of G_1 which were determined experimentally. Such values of G_4 and the corresponding values of $\int_0^f G_4df$ are given in Table XXX and the integral is given in a more extended manner in Table IX of the main paper.

The values of β_0', the threshold values while listening to speech with no noise present, can also be obtained from the values of G_1 by Eq. (94). Since $Z_0=B_s+\Delta B_s+\kappa-\beta_0'$, the values are given by the equation

$$\beta_0'=B_s+\Delta B_s-10\log G_1-\text{constant}. \quad (98)$$

The spectrum level for speech B_s and peak levels ΔB_s[23] and G_1 are given in Table XXX. This spectrum level is for a speaker having a talking level 68 db and whose spectrum level curve has the shape adopted as typical by French and Steinberg.[3] The constant was taken equal to 70 db so that values of β_0' and β_0 would agree in the frequency range from 1250 to 1750 c.p.s. The values thus calculated are given in column 8 of Table XXX. It will be seen from Fig. 50 that values of G_1 above 2000 cannot be considered very accurate so for the frequencies above this an average between β_0 for pure tone and the quantity $(B_s+\Delta B_s-10\log G_1-70)$ was taken for final values of β_0'. These values are tabulated in column 10. The values of κ, the critical band widths in db for monaural listening, are given in column 11.[3] The values of $\kappa-\beta_0'$ which are used in

[23] H. K. Dunn and S. D. White, "Statistical measurements on conversational speech," J. Acous. Soc. Am. 11, 278 (1940).

noise calculations are given in column 12, and values of Z_F from Eq. (41) or

$$Z_F = B_s + \Delta B_s + 8 + \kappa - \beta_0' \quad (41)$$

are given in the last column.

APPENDIX 2: HEARING LOSS FOR SPEECH

In this appendix the relationship between the hearing loss for speech[24] and the hearing loss audiogram will be considered. Let β_f be the hearing loss at the frequency f for a pure tone. It is the ordinate in the audiogram. If we consider β_f has the same effect upon the threshold level as an attenuation $-R$ from the flat response system, then by analogy to Eq. (23) the hearing loss for speech β_s is given by

$$10^{-\beta_s/10} = \int_0^\infty G_1 10^{-\beta_f/10} df. \quad (99)$$

[24] H. Fletcher, "A method of calculating hearing loss for speech from an audiogram," J. Acous. Soc. Am. 22, 1 (1950).

If we consider only the octave frequencies 125, 250, 500, 1000, 2000, 4000, and 8000, then the following equation is approximately correct.

$$\beta_s = -10 \log\{\sum W_k 10^{-\beta_k/10}\} \quad (100)$$

where k takes the successive values of 125, 250, 500, 1000, 2000, 4000, and 8000. The weights are given by

$$W_k = \int_{0.7k}^{1.4k} G_1 df.$$

For 125 c.p.s. $W=0.000$, 250 c.p.s. $W=0.003$, 500 c.p.s. $W=0.104$, 1000 c.p.s. $W=0.388$, 2000 c.p.s. $W=0.395$, 4000 c.p.s. $W=0.106$, 8000 c.p.s. $W=0.004$. So for most purposes one needs to consider only the four frequencies 500, 1000, 2000, and 4000 and use weights 0.1, 0.4, 0.4 and 0.1.

For a fairly flat audiogram it is approximately correct to take an average of the hearing loss at 500, 1000, and 2000 c.p.s.

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOLUME 22, NUMBER 2 MARCH, 1950

The Speaking Machine of Wolfgang von Kempelen*

HOMER DUDLEY
Bell Telephone Laboratories, Murray Hill, New Jersey

AND

T. H. TARNOCZY
Biophysical Laboratory of the Institute for Anthropology of the Museum of Natural History, Budapest, Hungary
(Received October 17, 1949)

The physiological motions involved in speaking can be indicated to the eye or to the ear. For the eye suitably chosen symbols may be written to indicate the physiological positions assumed in forming each sound; for the ear synthetic sounds may be produced by motions in a mechanism built to simulate the speech organs. The degree of phonetic success may be estimated in the case of the visible symbols by listening to sounds formed when the indicated physiological processes are carried out, and in the case of the speech-simulating mechanism by comparing the synthetic speech produced to normally spoken speech. Significant advances along both the visual and the aural lines are described from earliest times down to the present.

Wolfgang von Kempelen produced the first speaking machine worthy of the name around 1780. This paper gives his background, a description of the apparatus he built, and a discussion of the methods used in producing the various sounds, fitting his work into the over-all picture of speech-imitating devices from the speaking of idols of ancient times down to the automatic electrical reconstructing of speech in the vocoder. For portraying to the eye the physiological characteristics of speech there are discussed the more outstanding methods from claimed symbolic alphabets of ancient languages down to the recent spectrographic visible speech.

TOWARD the end of the 18th Century a Hungarian, Wolfgang Ritter von Kempelen, or, in Hungarian, Kempelen Farkas Lovag, first built a complete and, on the whole, a surprisingly successful speaking machine. Speech was formed by manipulation of mechanical elements simulating the essential parts of the human vocal system. In 1791 he published a 456-page book,[1] illustrated with 25 plates, describing his observations on human speech production and his experiments during the two decades he had been working on his speaking machine. The appearance of his book was a great social event. Introductory to the

* Orally presented before the Acoustical Society of America, May 5, 1949, New York, by Dudley with original draft by Tarnoczy. The paper here, in general, follows the oral presentation including a set of figures and also other material not in the original draft.

[1] *Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine.* Also published in French at the same time (1791).