An Appraisal of Received Telephone Speech Volume

By O. H. COOLIDGE and G. C. REIER

(Manuscript received November 3, 1958)

One of the attributes of telephone service which is of importance to a telephone user is the loudness with which he hears the voice of a distant talker. Related to this loudness is an objective measurement of "received volume."

This paper represents the results of subjective tests made to determine a relationship between received volume and the satisfaction of telephone listeners. The results are shown as statistical distributions of listeners' opinion, which may be combined with estimated distributions of received volume in the telephone plant to give "grade of service."

Grade of service objectives have been stated as 95 per cent of connections rated "Good", 5 per cent "Fair" and a negligible percentage "Poor." Increased use of the more efficient 500-type telephone sets, and planned improvements in the circuits which interconnect them, will make it possible to meet these objectives.

I. INTRODUCTION

Outside of his monthly telephone bills, an occasional visit to the Telephone Company commercial office and now and then a brief telephone conversation with the telephone operator, the customer's only contact with the telephone system is his telephone set. He knows that the wires and cables he sees strung on poles and the telephone building he occasionally passes on his way downtown bear some relation to the telephone in his home or office, but this is not very important to him. The important thing is his telephone set. Through it he can communicate with anyone else who has access to a telephone, no matter how far away that person happens to be.

Now what does the user expect from his telephone? We might say that he wants it to be reasonably pleasing in appearance, comfortable to use and simple to operate. He expects accuracy; that is, he wants to be connected to the party he calls and not to some stranger, and he does not want to be annoyed by answering "wrong numbers". Our typical customer is impatient when his call is unduly delayed without an explanation as to the reason for the delay. He wants a telephone bell or other signal which can be heard in all parts of his home but which is not objectionably loud if he happens to be near the telephone when it rings. He will object to "clicks" or other loud or unpleasant noises from his telephone. He is annoyed if he can hear "crosstalk" which is intelligible or nearly intelligible, because it implies violation of the privacy he expects in his own telephone conversation. He wants to hear in his telephone receiver a reasonably faithful, undistorted reproduction of the voice of the speaker. Finally, the most important attribute a customer expects of his telephone set — important because, without it, he might find no reason to have a telephone at all — is that it permit him to hear and to be heard with a minimum of effort on his part.

It is with this last attribute that the present paper is concerned. Briefly, if a customer hears the voice of his partner in telephone conversation with sufficient loudness and freedom from distortion and noise, little effort will be required to hear and understand what is said. If the distant voice is weak or distorted, the strain of listening requires effort. Likewise, if the distant party hears only with effort he, in turn, may request our typical customer to talk more loudly, which again calls for effort. In telephone systems of modern design, distortion and noise are no longer troublesome factors, so we shall consider only the loudness aspect. Thus, this paper is concerned with the determination of the magnitudes of received volume, or loudness, which give varying degrees of satisfaction to users, and with obtaining an estimate of how well the grade of transmission service provided by the Bell System meets the objective of satisfying customers in this respect.

What the customer would like to have and what it is economically feasible to give him may not be entirely compatible. Giving him what he wants has not always been feasible in the past, but, with the more general application of the improved telephone facilities that have become available over the past decade, this now appears within the realm of possibility.

What do we mean by "what he wants"? We must bear in mind that people vary widely in their judgment of preferred loudness, just as they do in their judgments involving their senses of feeling, sight, taste or smell. This is therefore a statistical problem involving the likes and dislikes of a large number of people. Their combined judgment might be expected to approximate (except at the tails) some distribution related to the normal law. A small number of telephone users will be quite tolerant of low received volumes, and another small number at the opposite end of the distribution will desire very loud volumes. But the large bulk of the telephone users will prefer some loudness level about midway between these extremes. It is the job of the engineer to find out by experiment that area of received volumes wherein satisfaction for the greatest number lies.

To carry out such an experiment with the telephone-using public would be a prodigious and impractical undertaking. However, by sampling methods and by closely controlled experimentation one can obtain an answer in the laboratory. The first thing to be done is to decide on a yardstick for expressing degree of satisfaction. Testing procedures are then set up whereby people can listen to different grades of transmission and then express their opinions in terms of the selected yardstick. Measurements of this type are known as "subjective" measurements. The results obtained depend wholly on the judgments of those taking the test. For this reason, large numbers of participants are required to obtain reliable answers.

Laboratory tests of this nature were conducted at Bell Telephone Laboratories over different portions of the total range of received telephone speech volumes at various times from 1947 to 1954. The results of all these tests are combined and summarized here.

II. RESULTS

Two types of subjective transmission tests commonly made in the laboratory are here called "appraisal tests" and "comparison tests". During the appraisal tests observers listen to speech over a telephone connection with specified transmission parameters such as volume, noise interference, etc. The observers are then asked to give their opinions of the transmission qualities of the connection by assigning it to one of several specified categories such as "Good", "Fair", "Poor", or the like. One of the transmission parameters is then changed and the process repeated. Each condition is thus rated on its own merits without direct comparison to any reference condition. In comparison tests, on the other hand, each transmission condition is compared immediately with some other condition, usually a known reference condition, and observers are asked which of the two they prefer in each case. The two types of subjective tests are useful under different circumstances that need not be pursued further here. The point to be noted is that results shown in this paper, except in one case, are based on appraisal tests.

Distributions of appraisal categories in a very wide range of received telephone speech volumes are shown in Fig. 1. The range shown is con-

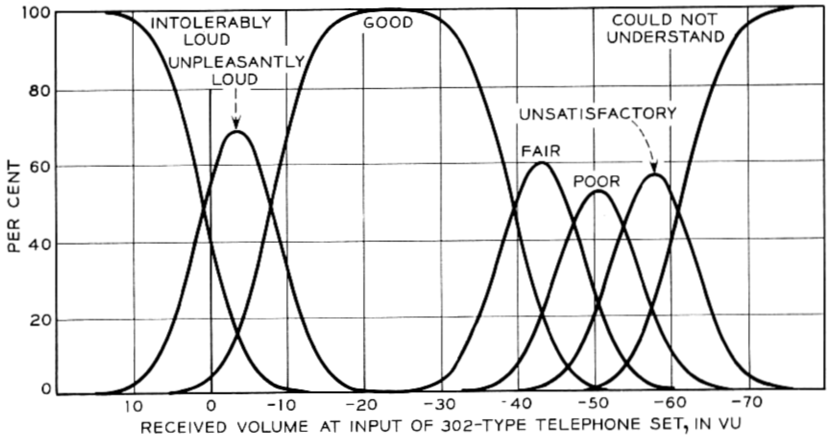


Fig. 1 — Opinion distributions — percentage of observations in which various values of received volume are assigned to the indicated basic categories.

siderably greater than any to be expected in commercial telephone service, since it extends from received volumes which are unbearably loud to those which are practically inaudible. Within this range there are seven natural categories descriptive of the transmission performance of a telephone connection at different levels of received volume in units known as VU. They are here called basic categories, and are as follows:

Intolerably Loud,
Unpleasantly Loud,
Good,
Fair,
Poor,
Unsatisfactory,
Could Not Understand.

The distribution curves for these categories must not be confused with probability density curves, under which the total area must always equal unity. Instead, each one shows the percentage of total observations at each volume level in which that level is assigned to the indicated category. Since any volume level must be assigned to *some* category, the sum of the percentages of the various categories at any one level must be 100. For example, for a received volume of −10 VU at the line terminals of a 302-type telephone set, 33 per cent of the listeners would say the volume is "Unpleasantly Loud" and 67 per cent would say that it is "Good". For a received volume of −25 VU, everybody would rate the call as "Good". For a received volume of −40 VU, 43 per cent would vote "Good", 50 per cent "Fair" and 7 per cent "Poor".

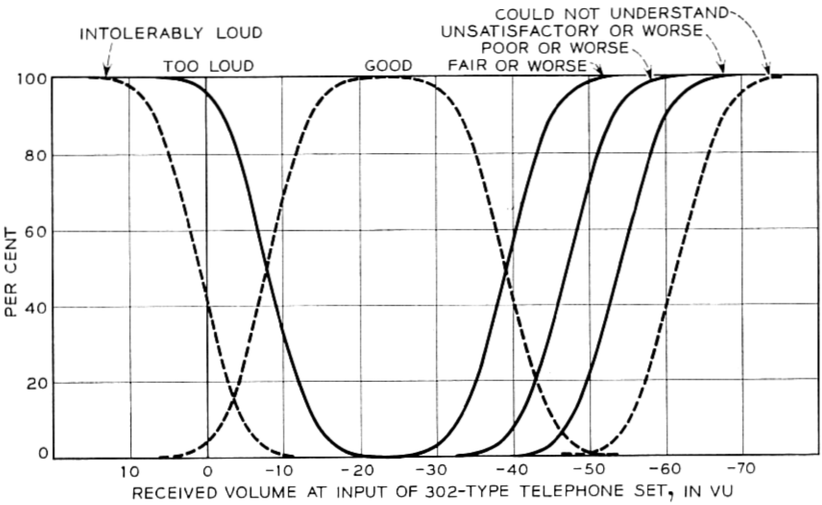


Fig. 2 — Opinion distributions — percentage of observations in which various values of received volume are assigned to the indicated categories. Solid curves are cumulative categories; dashed curves are basic categories.

The various appraisal categories in Fig. 1 are centered about the "Good" category. Curves to the left of "Good" indicate categories which are progressively less acceptable because of undesirably higher received volumes. The categories to the right are progressively less acceptable because of lower and lower received volumes. The two end categories, "Intolerably Loud" and "Could Not Understand" act as barriers because there is no further category on one side of each of them. Thus, each must eventually reach 100 per cent at sufficiently high (or low) volume levels. The distribution of the "Good" category also reaches a maximum of 100 per cent, not because it is a barrier, but because it covers such a wide range of received volumes that there is practically unanimous opinion in the middle of that range, which is around -23 or -24 VU.

The data shown in Fig. 1 can also be plotted in cumulative categories, as shown by the solid curves in Fig. 2. The three dashed curves, "Intolerably Loud", "Good" and "Could Not Understand" are individual (basic) rather than cumulative categories, copied from Fig. 1 for purposes of orientation. The cumulative categories at the high-volume end include all individual categories to their left, while the cumulative categories at the low-volume end include all individual categories to their right, as may be seen by comparing Figs. 1 and 2. Thus, the percentage value at any volume level for "Too Loud" on Fig. 2 is the sum of the percentage values for "Unpleasantly Loud" and "Intolerably Loud" on

Fig. 1. For example, at −5 VU, 67 per cent of the listeners vote "Unpleasantly Loud" and 8 per cent vote "Intolerably Loud". The percentage in the cumulative category "Too Loud" is therefore 75, as shown in Fig. 2. Likewise, at −45 VU 56 per cent vote "Fair", 30 per cent "Poor" and 3 per cent "Unsatisfactory". Thus, on Fig. 2 "Fair or Worse" is 89 per cent and "Poor or Worse" is 33 per cent. It will be noticed that each of the cumulative categories in turn takes over the functions of "Intolerably Loud" or "Could Not Understand" in acting as an ultimate barrier, and thus eventually reaches a value of 100 per cent.

It would be too much to expect reliable results from tests at one sitting, if observers were required to carry in their heads the large number of basic categories shown by the curves of Figs. 1 and 2. Consequently, the actual tests were made over reduced ranges of received volume, one series in the lower range embracing the five categories from "Good" to "Could Not Understand", inclusive, and another series in the upper range including the three categories from "Good" through "Intolerably Loud".

One is struck also by the tolerance implied by the great width of the "Good" distribution. In addition to being a fortunate fact, it suggests that there must be some smaller range of volumes included within the "Good" category which would be found to be most acceptable of all. This preferred range, or "Excellent" category as it has sometimes been called, has been determined by two methods and is shown in Fig. 3. It was originally determined by appraisal tests similar to those employed in determining the results shown in Figs. 1 and 2. Some years later, the earlier determination was checked by a different technique employing comparison tests. In the comparison tests ten different volume levels (covering the "Good" range of Fig. 1) were compared, each with every other level. Each volume level was then rated according to the percentage of times it was preferred over the other levels with which it was compared.

Fig. 3 shows that the two methods give results which check each other satisfactorily, since the modal points differ by only 1 db. Using the appraisal distribution (to be consistent with the major tests on Figs. 1 and 2), it appears that a received volume level of −19 VU (the modal point of the curve) is the value preferred over all others. It is interesting to look back at Fig. 1 and note that −19 VU occurs, not at the center of the "Good" distribution, but well toward the high-volume side of its table top. This indicates that listeners prefer about the highest volume they can get just short of the point where too much loudness becomes annoying.

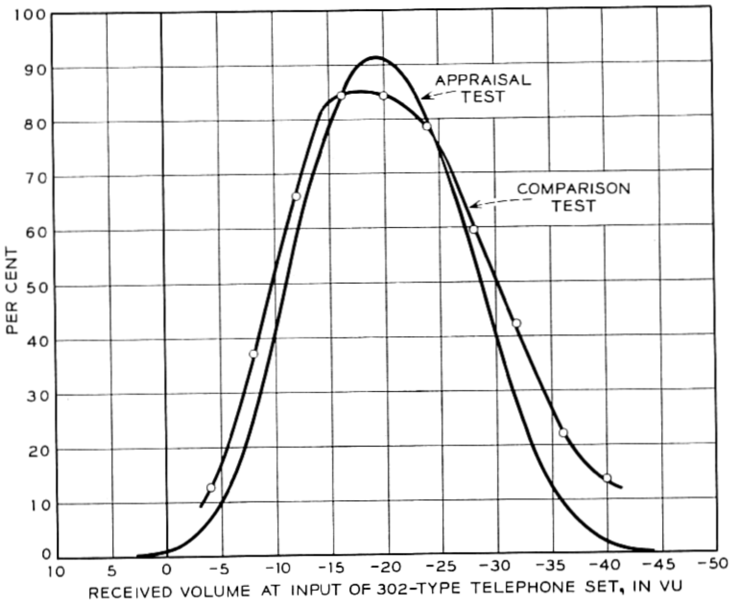


Fig. 3 — Preferred level of received volume, appraisal test versus comparison test. Appraisal test curve shows percentage of observations in which the indicated volume level is assigned to the preferred category; modal point is −19 VU. Comparison test curve shows percentage of observations in which the indicated level is preferred in comparison with other levels tested; modal point is −18 VU.

III. DESCRIPTION OF TESTS

It was recognized that the intelligibility of received telephone speech, and the listener's satisfaction, are affected by transmission factors other than just speech volume, or loudness. Consequently, wherever the effect of other factors might have been significant, they were controlled during the tests at values which cause little or no transmission impairment. The factors which were controlled and their values are as follows:

line noise: 17 dba at telephone receiver;*
room noise: 50 db RAP (reference acoustic pressure.);
speech transmission band: 150 to 3200 cps.

The test circuit is shown schematically in Fig. 4. It is patterned on a typical telephone circuit employing 302-type telephone sets, but with means provided for varying the loss of the trunk so that different levels

* Dba is the unit employed in the Bell System for measurements of line noise with the 2B Noise Measuring Set. This unit was adopted when the 302-type telephone set came into use, in order to provide equal numerical readings of the 2B set when noise of equal impairment was encountered in circuits with different types of telephone set.

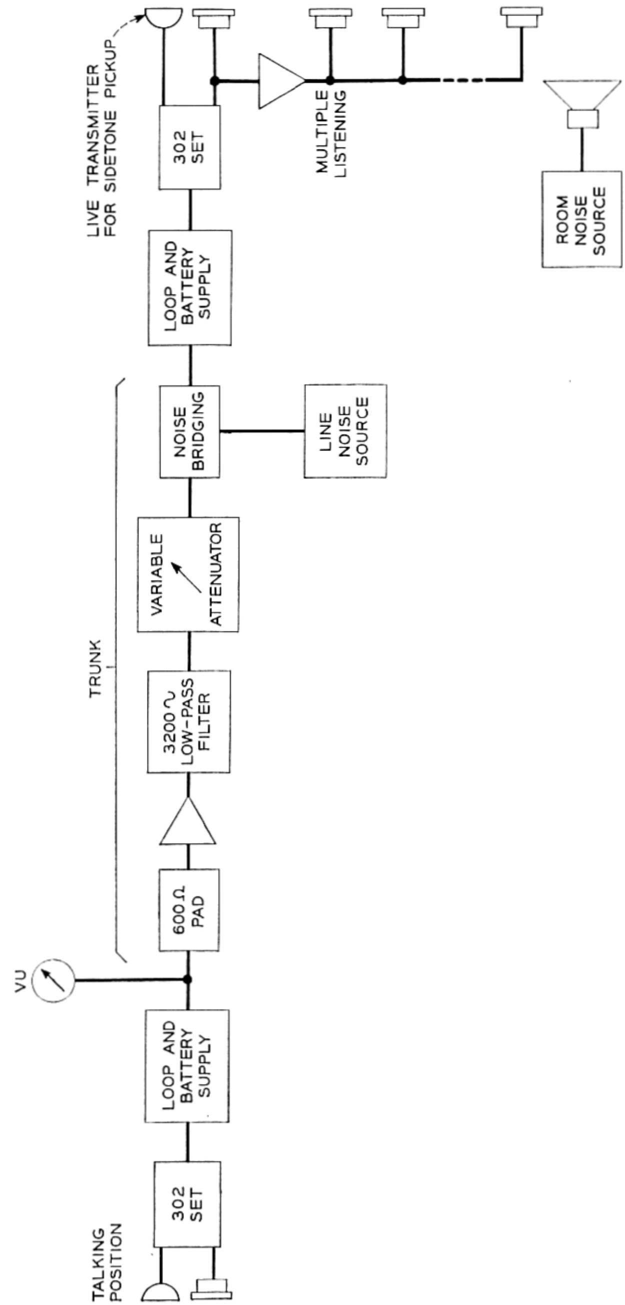


Fig. 4 — Circuit used for appraisal tests.

Fig. 5 — Observers at multiple listening positions during appraisal tests.

of speech volume may be obtained at the receiving end. A volume indicator at the transmitting end of the test circuit enables the talker to keep the level of his own speech constant, thus providing means for determining, from the known losses in between, the level of speech volume at the receiving telephone. The transmitter of the receiving telephone set is kept operative, so that a normal amount of room noise will be present in the receiver through the sidetone path. A multiple listening arrangement is provided so that many observers may listen simultaneously, each hearing the same speech volume and room noise through the sidetone path that are present at the main receiving station. This permits gathering a large amount of data with a minimum expenditure of time and effort on the part of those conducting the tests. A group of observers is shown taking the test at the multiple positions in the photograph in Fig. 5. Fig. 6 shows the board at which the test circuit is set up and controlled, and the main receiving station to which the observers' handsets are multipled. The talking position does not show in the photographs since it is located in an adjacent soundproofed room.

At first, three men and one woman were used as talkers. It was found that differences in the type of voice (even in the case of the woman talker)

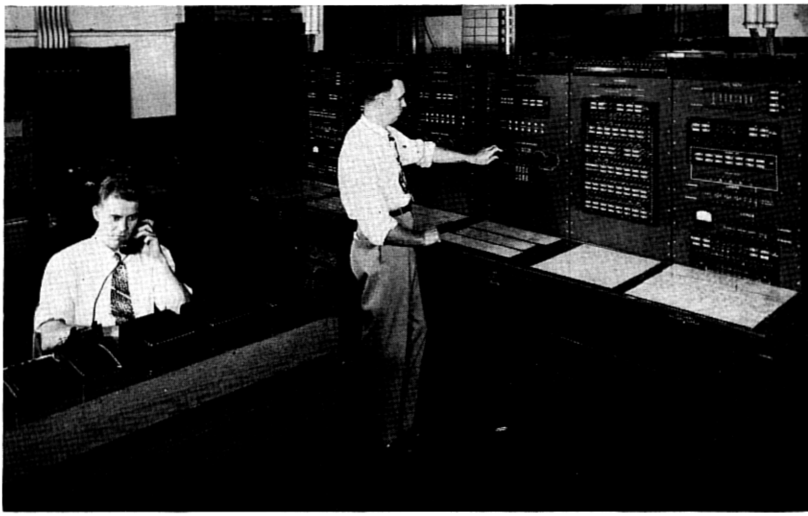


Fig. 6 — Circuit control board and main receiving station for appraisal tests.

had almost no effect on the opinions of listeners, provided that each talker impressed the same speech volume on the line. Consequently, in the later tests almost all of the talking was done by one of the original male talkers.

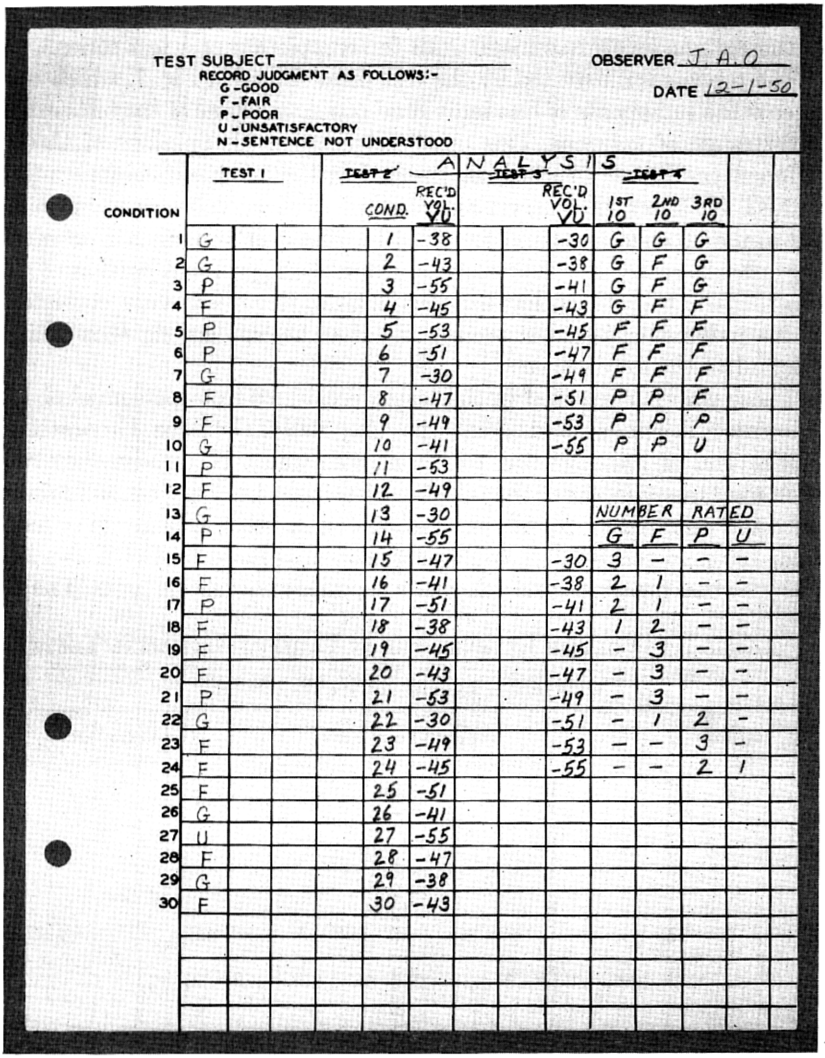
The test itself was a very simple and straightforward procedure. Observers were allowed to listen to speech over the test circuit at a particular level of received volume. They were then asked to assign that sample to one of the categories previously listed for them. In the case of the tests at the lower levels the categories, as already stated, were "Good", "Fair", "Poor", "Unsatisfactory" and "Could Not Understand", while those in the high-level tests were "Good", "Unpleasantly Loud" and "Intolerably Loud". No attempt was made to define these categories for the observers, each one deciding for himself what the terms meant. When observers had recorded their opinions of the first speech sample, the received volume was changed by adjusting the variable attenuator shown in Fig. 4, and the process was repeated until the entire range had been covered three times. The specific values of received volume used and the ranges covered in the different series of appraisal tests are shown in Table I. The different volume levels were presented to listeners in random order. This is necessary because it has been found that sequential orders (high to low or low to high) result in displacements of the entire opinion distribution, apparently due to conditioning of observers by volume levels to which they have become accustomed. No disclosure of the magnitudes of the volume levels was made until the completion of a test series.

The process is illustrated by the test sheet shown in Fig. 7, which represents the judgments of one individual observer in one of the tests in the lower range of volumes. This observer recorded his opinions of the 30 different conditions to which he listened in the left-hand column under "TEST 1". Later, the analyst recorded the received volumes corresponding to the 30 conditions and arranged the observer's opinions in descending order of received volumes, as shown in the right-hand portion of the sheet. It will be noticed that this particular observer was very consistent in his judgments (as most were), and that he was among those fairly tolerant of the lowest levels.

Those who participated in the tests were drawn from personnel of the American Telephone and Telegraph Co., Bell Telephone Laboratories and several of the operating telephone companies. Naturally, they represented many facets of the telephone business, but only a small minority could be rated as experts in transmission matters. One group of more than 150 men consisted of young engineers just hired by Bell Laboratories who had practically no telephone experience except as users. The re-

TABLE I — VOLUME LEVELS USED IN VARIOUS APPRAISAL TESTS

VU at Input of Receiving 302-Type Telephone Set		
High-Range Tests	Preferred-Range Tests	Low-Range Tests
+6		
+3	+2	
0		
−3		
−6	−6	
−9	−9	
−12	−12	
−15	−15	
−18	−18	
−21	−21	
	−24	
	−27	
	−30	−30
	−33	
	−36	
	−39	−38
		−41
		−43
		−45
	−48	−47
		−49
		−51
		−53
		−55



TEST SUBJECT ______ OBSERVER J. A. O.

RECORD JUDGMENT AS FOLLOWS:-
G - GOOD
F - FAIR
P - POOR
U - UNSATISFACTORY
N - SENTENCE NOT UNDERSTOOD

DATE 12-1-50

ANALYSIS

CONDITION	TEST 1	~~TEST 2~~ COND.	REC'D VOL. VU	~~TEST 3~~ REC'D VOL. VU	~~TEST 4~~ 1ST 10	2ND 10	3RD 10	
1	G	1	-38	-30	G	G	G	
2	G	2	-43	-38	G	F	G	
3	P	3	-55	-41	G	F	G	
4	F	4	-45	-43	G	F	F	
5	P	5	-53	-45	F	F	F	
6	P	6	-51	-47	F	F	F	
7	G	7	-30	-49	F	F	F	
8	F	8	-47	-51	P	P	F	
9	F	9	-49	-53	P	P	P	
10	G	10	-41	-55	P	P	U	
11	P	11	-53					
12	F	12	-49					
13	G	13	-30		NUMBER	RATED		
14	P	14	-55		G	F	P	U
15	F	15	-47	-30	3	–	–	–
16	F	16	-41	-38	2	1	–	–
17	P	17	-51	-41	2	1	–	–
18	F	18	-38	-43	1	2	–	–
19	F	19	-45	-45	–	3	–	–
20	F	20	-43	-47	–	3	–	–
21	P	21	-53	-49	–	3	–	–
22	G	22	-30	-51	–	1	2	–
23	F	23	-49	-53	–	–	3	–
24	F	24	-45	-55	–	–	2	1
25	F	25	-51					
26	G	26	-41					
27	U	27	-55					
28	F	28	-47					
29	G	29	-38					
30	F	30	-43					

Fig. 7 — Sample test sheet for appraisal tests.

sults obtained from these various groups were analyzed separately to determine whether different age groups (with corresponding differences in hearing acuity) or differences in transmission background might be a factor in their appraisals of received volume. In general, it was found that there was no significant difference between the opinions of different groups. For example, the data obtained from the student engineer group alone differed from the data for the entire group by no more than ± 0.5 db.

A sample of the data in cumulative categories is shown in Fig. 8. By plotting the experimental points on arithmetic-probability paper one is able to smooth the data by passing a straight line through the data points. Except for some divergence at the tails of the distributions, the data are a reasonable approximation to the straight lines, and therefore to the normal law of error. The distributions of cumulative categories in Fig. 2 are taken directly from the straight lines of Fig. 8 and from similar distributions for the categories not shown here. The distributions of basic categories in Fig. 1 are obtained by taking vertical differences between the straight lines of Fig. 8 at each level of received volume. Thus, "Unpleasantly Loud" is the difference between "Too Loud" and "Intolerably

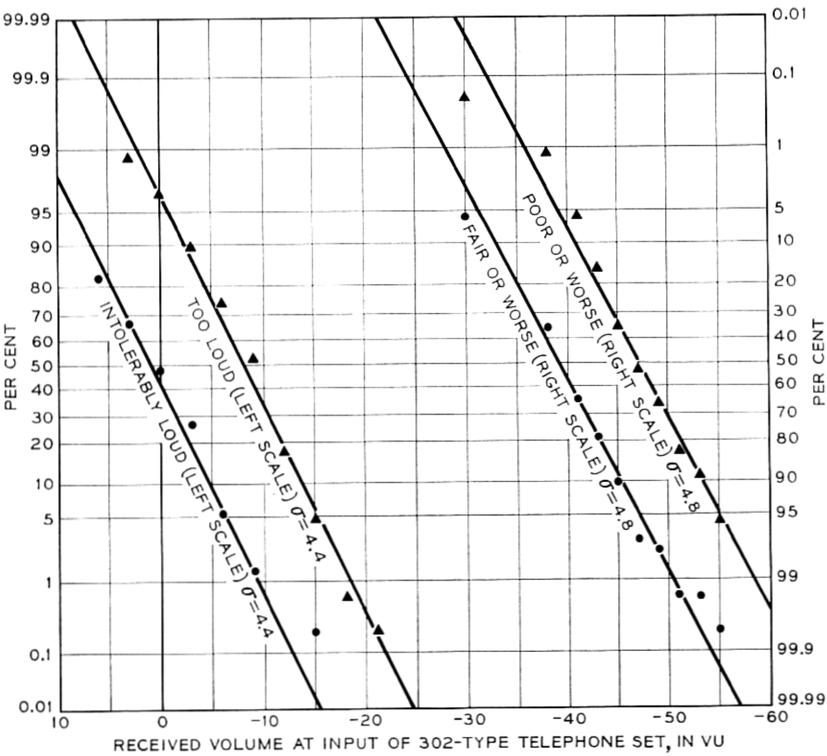


Fig. 8 — Sample of data for opinion distributions — percentage of observations in which various received volumes are assigned to indicated cumulative categories.

Loud", "Fair" is the difference between "Fair or Worse" and "Poor or Worse", and so on.

The reliability of the data is attested by the fact that, for the low-volume categories, they are based on approximately 1400 observations by 470 different listeners at each of the values of received volume indicated in Table I. The data for the high-volume categories are the result of 528 observations by 176 observers at each indicated level.

These two series have been spoken of here as the "major" tests to distinguish them from the tests of the preferred range of received volume levels, the results of which are shown in Fig. 3. The appraisal tests of Fig. 3 were made in the same manner as the major tests (at an earlier date), but were based on fewer observations, 56 at each of the levels indicated on Table I. It was decided that results based on so few observations might be questionable, and that further tests to confirm the earlier results were warranted. Instead of additional appraisal tests, the type of comparison tests which has already been described seemed applicable to the objective of obtaining the preferred volume level, and was adopted.

The talking circuit for the comparison tests was the same as that used in the appraisal tests (Fig. 4). In these tests each of the ten selected volume levels was compared with every other level twice, once with the level A preceding level B in the presentation, and again in the opposite order, B preceding A. However, the two comparisons of like levels were not made consecutively, but were interspersed among other combinations. The comparisons were thus presented to observers in an order that was random both as to the volume levels involved in the comparison and

TABLE II — ANALYSIS OF PREFERRED-RANGE COMPARISON TESTS

Level Preferred	Times That Volume Level in Left Column Was Preferred Over Level in Headings Below										W	L	%
	−4	−8	−12	−16	−20	−24	−28	−32	−36	−40			
−4	—	1	0	0	0	0	0	2	7	6	16	110	12.7
−8	13	—	0	0	1	4	5	6	7	11	47	79	37.3
−12	14	14	—	1	5	7	7	9	14	12	83	43	65.8
−16	14	14	13	—	7	7	13	13	13	12	106	20	84.2
−20	14	13	9	7	—	8	13	14	14	14	106	20	84.2
−24	14	10	7	7	6	—	13	14	14	14	99	27	78.5
−28	14	9	7	1	1	1	—	14	14	14	75	51	59.5
−32	12	8	5	1	0	0	0	—	14	13	53	73	42.1
−36	7	7	0	1	0	0	0	0	—	13	28	98	22.2
−40	8	3	2	2	0	0	0	1	1	—	17	109	13.5
	110	79	43	20	20	27	51	73	98	109			

W (Won) denotes that the level in left column was preferred.
L (Lost) denotes that the level in left column was not preferred.

as to the precedence of one value over the other in a combination. As in the appraisal tests, the magnitudes of received volumes were not disclosed to observers until after the tests were completed.

The method of analysis of the comparison test data is somewhat unusual. If each observed comparison between two volume levels is considered as a game between two teams, the "games won" and "games lost" may be charted in the same manner as the results usually published for baseball leagues for a season in which each team plays the same number of games with every other team. This analysis is shown in Table II, the right-hand column of which shows the percentage of games "won" by each of the ten volume levels, that is, the percentage of observed comparisons in which each of the indicated levels is preferred over the other nine levels with which it is compared. The values from this column are plotted as data points for the comparison test distribution shown in Fig. 3.

IV. APPLICATIONS

The tests which have been described and the curves which show the results of these tests provide fundamental data which, in the authors' opinion, are indicative of a telephone customer's expectancy with regard to hearing and being heard with a minimum of effort on his part. In this respect, they are a measure of what is required to satisfy telephone users. By themselves, the results of the tests indicate observers' opinions (as to category) of any specific value of received volume, such as might be encountered on any one individual telephone connection.

While this is of interest, it is more important to the management of an operating telephone company to know customers' reactions to the grade of service provided over that company's telephone plant, since this information may affect decisions on such questions as spending money for plant improvements. Specifically, the management should know the percentage of telephone connections over the plant which customers consider good, the percentage fair, the percentage poor, etc. Such information may be obtained by combining the opinion distributions described here with the estimated distribution of received volumes which customers actually obtain in their daily use of the telephone. This is done by integrating, over the range of received volumes, the compound probability: (a) that the telephone user considers a particular volume "Good" (or "Fair" or "Poor"), and (b) that he actually receives that volume. It should be pointed out that the distribution of volumes actually received is, in turn, a combination of the distribution of talking volumes, which vary over a range of some 30 VU, and the distribution of plant losses from the point of the talking volume measurement to the input of a listener's telephone

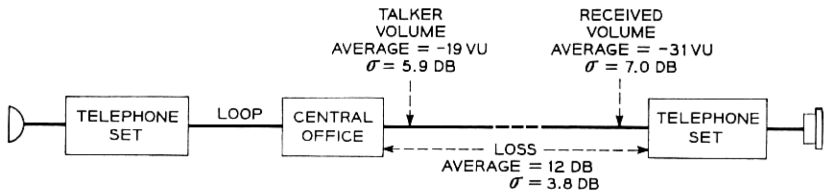


Fig. 9 — Circuit showing points of volume measurement and intervening plant losses.

set. Fig. 9 is a simple diagram which illustrates this point. It shows a transmitting telephone and a loop to the central office. Field measurements of telephone users' talking volume indicated that the average volume input into the line leaving the central office was −19 VU, with a standard deviation of 5.9 db at the time the survey was made.[1] An analysis made several years ago of the large variety of telephone connections encountered in the telephone plant indicated that the average transmission loss between the talker's central office and the distant listener's telephone set was 12 db, with a standard deviation of 3.8 db. The average received volume at the telephone terminals was therefore −31 VU, with a standard deviation of 7.0 db.*

The results of combining the distributions of observers' opinions and actual received volumes have been called "grade of service". "Grade of service" curves may be plotted showing the percentage of telephone connections considered "Good", "Fair", "Poor", etc., against the *average* of various received volume distributions, assuming that these distributions are allowed to vary in average value, but not in standard deviation. In this form they have been found useful in setting objectives for received volumes to be provided and for plant losses which will permit attainment of those received volumes. An example of grade of service curves in cumulative categories is shown in Fig. 10. The same data in another form, which includes individual categories, appear in Fig. 11. For an over-all picture of grade of service provided in the plant just described, assuming that all telephone sets were of the 302 type, the average received volume may be taken as −31 VU, with a standard deviation of 7.0 db, as indicated in Fig. 9. This average value constitutes one point on the abscissae of Figs. 10 and 11. It will be noted in Fig. 11 that, if the average of the distribution of received volumes is −31 VU, the tele-

* In this calculation it is assumed that talker volume and plant loss are independent variables. It is recognized that there may be some small correlation between them. For instance, Subrizi[1] found a small correlation between talker volume and distance between talker and listener on very long distance calls. Other tests to determine correlation have given conflicting results so the correlation is probably small enough to justify the assumption.

−28 VU, rather than at the preferred volume level, −23 VU (for the future plant with 500 sets.) Once we have increased the average value of received volumes to this objective, no further improvement in grade of service may be expected (should it become desirable) except by lowering the standard deviation still further.

V. ACKNOWLEDGMENT

The authors wish to express their gratitude to C. W. Carter of Bell Telephone Laboratories for making available to them his 1947–48 data covering appraisal tests in the preferred range of received volumes. They also extend their thanks to H. R. Huntley, now Chief Engineer of the American Telephone and Telegraph Company; to W. E. Bloecker and the late L. B. Bogan of Mr. Huntley's former department for their support of this program; and, finally, to the many others in the American Telephone and Telegraph Company and Associated Companies, and in Bell Telephone Laboratories, for their conscientious efforts in forming the unbiased judgments which were so necessary for the success of these tests.

REFERENCES

1. Subrizi, V., A Speech Volume Survey on Telephone Message Circuits, Bell Labs. Record, **31**, August 1953, p. 292.
2. MacAdam, W. K., A Basis for Transmission Performance Objectives in a Telephone Communication System, Comm. & Elect., No. 36, May 1958, p. 205.

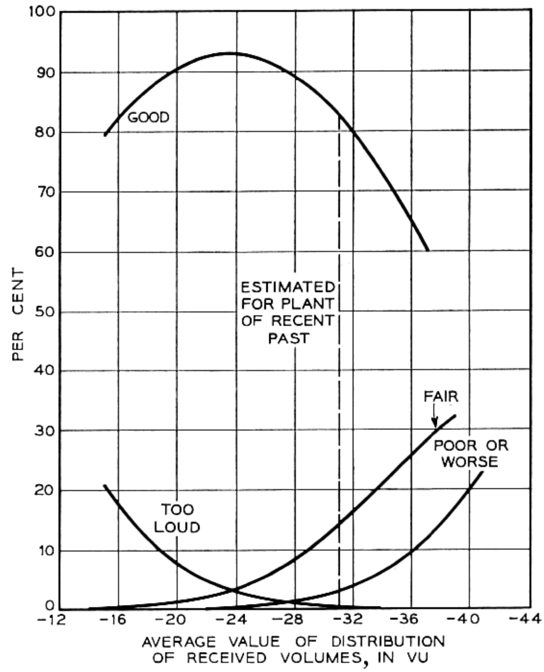


Fig. 11 — Grade of service distributions provided by plant of recent past — percentage of telephone connections assignable to various categories when the average of the received volume distribution has the indicated value.

is the same as that of the corresponding opinion curve on Fig. 8. By using arithmetic-probability paper, this simple construction of Fig. 10 replaces laborious summations of the compound probabilities. The summations cannot be avoided, however, if the distribution of received volumes is other than normal law.

For applications such as these, some reminders concerning the nature of the fundamental data presented here will not be out of place. Observers' opinions as to the category of specific values of received volume have been found to vary with (a) the amount of interference present in the form of line noise and ambient room noise at the listener's location, (b) the upper cutoff frequency of the telephone message channel and (c) the sensitivity of the listening telephone set and receiver, both for speech and for noise. As indicated early in this paper, data shown in Figs. 1 and 2 apply only to conditions of line noise, room noise and transmission bandwidth which have been found to cause little or no transmission impairment. Line noise and transmission bandwidth are controllable factors, and the values used are those which have been worked to as objectives in the Bell System for many years. Room noise is not under control of the telephone companies, but the value used is one that has been found representative of room noise in the average residence or fairly quiet office. In addition, the data of Figs. 1 and 2 apply only to receiving telephone sets of the 302 type, the type in greatest use in the telephone plant of the early 1950's, but now being largely supplanted by the 500 type. While all the conditions mentioned were reasonably normal a few years ago, it must be borne in mind that changes are being made over the years which necessitate adjustment of the opinion distributions with respect to the VU scale.

Long-range objectives for received loudness in the Bell System have been the subject of much study in recent years. W. K. MacAdam, Transmission Engineer of the American Telephone and Telegraph Company, has stated[2] that the design objectives of the Bell System might be about as follows:

i. A negligible number of calls rated "Poor".
ii. No more than 5 per cent rated "Fair".
iii. The balance rated "Good" or "Excellent".

It is evident that the telephone plant of a few years ago did not fully meet this objective. However, the picture is altered in the case of the future plant. If modern sets of the 500 type having higher receiving sensitivity are employed, the opinion curves will be found to shift with respect to the VU scale by approximately the amount of the sensitivity difference.* Thus, a value around −23 VU, rather than −19, would represent the volume level preferred over all others. In addition, it has been found that, with the 500-type telephone set, although the average talker volume output for a given level of acoustic input is increased, the spread in volume is decreased (smaller standard deviation) because there is a certain amount of speech compression on higher volumes. Furthermore, planned improvements in telephone lines (more precisely, the transmission medium between telephone sets) are expected to result in lower losses and lower standard deviation. It should be noted that, in the region of received volumes in which we are interested, lower standard deviation contributes to improved grade of service, just as higher average values of received volume do.

Based on these and other considerations, estimates have been made of the grade of service which will be provided by the future telephone plant, assuming that it includes the 500-type telephone set preponderantly,

* Assuming the total noise reaching the user's ear remains unchanged. Because of improvements in the sidetone circuit this is roughly the case.

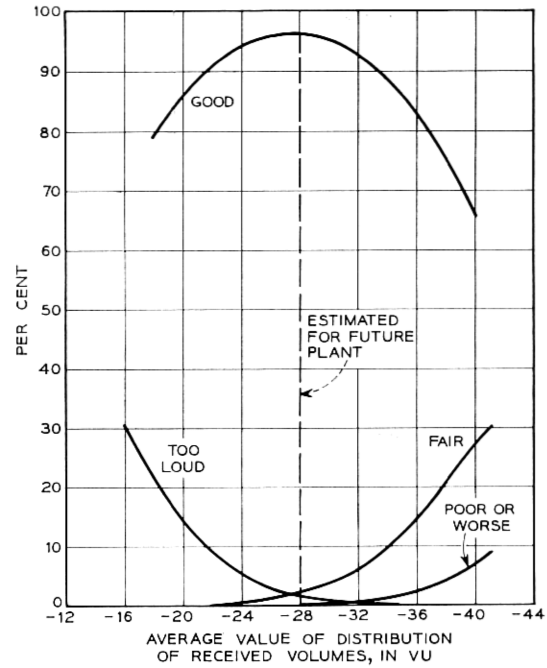


Fig. 12 — Grade of service distributions estimated for the future plant — percentage of telephone connections assignable to various categories when the average of the received volume distribution has the indicated value.

and the contemplated line improvements. Distribution curves based on these estimates are shown in Fig. 12. It is estimated that the average received volume resulting from the improvements mentioned will be −28 VU, as indicated in this figure. The percentage of calls rated "Good" would then exceed the objective of 95, while the calls rated "Fair" would come well within the 5 per cent objective. This is a satisfactory outlook for the future.

We then shall have gone about as far in the direction of increasing received volumes as we should. Fig. 12 shows that, at an average received volume of −28 VU, the percentage of calls considered "Too Loud" is about 1.5. Any further increase in received volume might raise the percentage "Too Loud" to undesirable values, with a corresponding decrease in the percentage of calls rated "Good". The fact that the "Good" distribution turns down at higher volume levels is the reason why it is advisable to set an objective for the average value of the received volume distribution at a volume level near the peak of the "Good" distribution, around

−28 VU, rather than at the preferred volume level, −23 VU (for the future plant with 500 sets.) Once we have increased the average value of received volumes to this objective, no further improvement in grade of service may be expected (should it become desirable) except by lowering the standard deviation still further.

V. ACKNOWLEDGMENT

The authors wish to express their gratitude to C. W. Carter of Bell Telephone Laboratories for making available to them his 1947–48 data covering appraisal tests in the preferred range of received volumes. They also extend their thanks to H. R. Huntley, now Chief Engineer of the American Telephone and Telegraph Company; to W. E. Bloecker and the late L. B. Bogan of Mr. Huntley's former department for their support of this program; and, finally, to the many others in the American Telephone and Telegraph Company and Associated Companies, and in Bell Telephone Laboratories, for their conscientious efforts in forming the unbiased judgments which were so necessary for the success of these tests.

REFERENCES

1. Subrizi, V., A Speech Volume Survey on Telephone Message Circuits, Bell Labs. Record, **31,** August 1953, p. 292.
2. MacAdam, W. K., A Basis for Transmission Performance Objectives in a Telephone Communication System, Comm. & Elect., No. 36, May 1958, p. 205.