Copyright © 1976 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 55, No. 9, November 1976
Printed in U.S.A.

Models for the Subjective Effects of Loss, Noise, and Talker Echo on Telephone Connections

By J. R. CAVANAUGH, R. W. HATCH, and J. L. SULLIVAN

(Manuscript received January 9, 1976)

Tests have been conducted at Bell Laboratories within the last 10 years to obtain subjective evaluations of the effects of loss, noise, and talker echo on telephone transmission quality. We use these subjective test results to formulate graphical and analytical models of subjective opinion that can be used in network planning studies to evaluate transmission performance of the network and to study the effects of network changes on performance. These models are based on the concept of a generalized transmission-rating scale. Separate opinion curves for each test take into account differences caused by factors such as subject group, type of test, and range of conditions. We also describe the methods of data analysis used in the formulation of the transmission-rating scale and opinion models, provide a comparison of the test results with the models, and discuss the models in sufficient detail to permit their application in transmission planning studies.

I. INTRODUCTION

In the last 10 years, several tests have been conducted at Bell Laboratories to determine the subjective evaluation of loss, noise, and talker echo in telephone connections. The purpose of these tests was to obtain information for use in network planning studies. Several hundred volunteers served as subjects and participated in several thousand test calls with various amounts of loss, noise, and talker echo. Some of these tests were conducted on normal business calls made by Bell Laboratories employees; others were conducted in a laboratory environment. At the end of each call, the subject was asked to indicate his or her opinion of the transmission quality on a five-category rating scale: excellent, good, fair, poor, and unsatisfactory.

The results from these tests were used to formulate the graphical and analytical models of opinion which are presented in this paper. These models have been used extensively in network planning studies to evaluate present network performance and to study the effect of possible changes in the network. Although many of these studies are not yet published, a recent paper by Spang[1] provides an excellent example of the application of these models in toll transmission planning.

For many years the Bell System has used subjective tests to obtain information on the effects of transmission quality which could be used in network planning studies. For example, the present Via Net Loss design plan, which introduces loss to control talker echo in the Direct Distance Dialing (DDD) network, is based on the results of subjective tests for talker echo that are included in a 1953 paper by Huntley.[2] Subsequent talker-echo tests were described by Phillips in 1954.[3] Coolidge and Reier reported the results of tests of received telephone speech volume in 1959 and introduced the concept of volume grade-of-service.[4] Test results for message circuit noise were used in noise grade-of-service studies described by Lewinski in 1964.[5]

Most of the tests mentioned considered the effect of one transmission parameter at a time. Since transmission parameters appear in combinations and there are, in many instances, important interactions, a new series of conversational tests for the combined effects of connection loss and circuit noise was initiated in 1965. Subjective evaluations were obtained on normal business calls within Bell Laboratories. The results of these tests were reported by Sen in 1971.[6] Since then, the use of combined loss-noise grade-of-service based on these tests has largely replaced the use of the earlier noise and volume grade-of-service.

Another test to determine subjective reaction to loss and noise was conducted in 1969 on normal business calls within the Bell Laboratories location at Holmdel, New Jersey. There were two reasons for the new test. One was to obtain a larger number of subjects per test condition and thus reduce the experimental variability. The other reason was to include both symmetric loss conditions, as was done in the 1965 tests, and asymmetric loss conditions; i.e., unequal loss in the two directions of transmission. The test results for the symmetric conditions indicated a more critical assessment of quality than the 1965 tests, which could not be explained by known differences in the tests. Therefore, a third test was planned and conducted in 1972.

During this same period of time, new talker-echo tests were initiated. These echo tests used the same five-category rating scale as the loss-noise tests so that possible tradeoffs between loss-noise and echo grade-of-service could be studied. Some of the echo tests were conducted in the laboratory. Others were conducted on normal business conversations between Bell Laboratories employees, where values of loss, noise, echo-path loss, and echo-path delay could all be controlled.

As the results of the various tests became available, work was continued to modify and improve earlier loss-noise opinion models and to develop similar talker-echo opinion models. Systematic methods were formulated to analyze the results from individual tests and to combine the results from different tests into a composite loss-noise-echo opinion model. In addition, a transmission-rating scale was introduced that assigned a single numerical value to any specific combination of transmission conditions.

The concept of a generalized transmission-rating scale recognized that subjective test results can be affected by various factors such as the subject group, the type of test, and the range of conditions that are included in the test. These factors were found to cause changes in both the mean opinion score for a given condition and in the standard deviation.* Thus, there were difficulties in trying to establish a unique relationship between a given transmission condition and subjective opinion in terms of mean opinion score or other subjective measures of transmission quality. The introduction of a transmission-rating scale tended to reduce this difficulty by separating the relationship between transmission characteristics and opinion ratings into two parts. For the first part, the transmission rating as a function of the transmission characteristic, was anchored for two specific transmission conditions and thus tended to be much less dependent on individual tests. The second part, the relationship between transmission rating and subjective opinion ratings, could then be displayed for the individual test.

The essential features of the transmission-rating scale and opinion models are summarized in Section II of this paper in enough detail to permit their application in transmission planning studies. The remainder of the paper describes three subjective tests for connection loudness loss and circuit noise and four tests for talker echo, outlines the methods of analysis, and describes the formulation of the composite transmission-rating scale and opinion models. Comparisons of the individual test results and the final model are also presented.

II. SUMMARY OF TRANSMISSION-RATING AND OPINION MODELS

The models for transmission rating described in this section are based on the results of seven subjective tests. All of the tests were conducted with Western Electric 500-type telephone sets.[7] Loudness

---

* Results of subjective tests in terms of the number of votes in each of the several categories of a rating scale can be expressed in a number of ways. One way is to assign numerical values to each of the categories, e.g., excellent = 5, good = 4, fair = 3, poor = 2, and unsatisfactory = 1. Each of these numerics is then weighted by the proportion of votes in the corresponding category for a particular transmission condition, and the weighted values summed. The result is called the mean opinion score for that transmission condition.

Table I—Summary of tests

	Loss-Noise Tests			Talker Echo Tests			
	SIBYL 1965 (MH)	SIBYL 1969 (HO1)	SIBYL 1972 (HO2)	Lab 1966	Lab 1968	Lab 1970	SIBYL 1970
Number of subjects	66	78	74	29	30	100	45
Number of conditions	24	15	12	30	93	10	16
Number of ratings	685	1163	1684	870	2790	1000	752
Median ratings/ condition	29	60	60	29	30	100	38
Connection loudness loss (dB)	5–30	10–30	5–30	*	*	18	10
Circuit noise (dBrnC)	21–44	22–45	25–42	28	18–38	33	30
Echo-path loudness loss (dB)	—	—	—	10–59	0–50	33–73	6–42
Echo-path delay (ms)	—	—	—	20–90	1.5–90	600, 1200	10–72
Sidetone-path loudness loss (dB)	13	12	12	9	12	12	12
Average room noise [dB(A)]	45	42	42	35	35	38	42

* These tests were not, strictly speaking, two-way conversation tests and, thus, connection loudness-loss values are not appropriate.

loss values used in the model describe the acoustic-to-acoustic transfer efficiency of overall telephone connections and are expressed in terms of the Electro-Acoustic Rating System (EARS) method.[8] Noise values used in the model are expressed at the line terminals of a telephone set with a reference receiving efficiency of 26 dB based on the EARS method.*

The major aspects of these tests are summarized in Table I. A more detailed description of the tests is presented in Sections III and IV.

The results of the subjective tests were used to derive transmission-rating models for (*i*) loss and noise, (*ii*) talker echo, and (*iii*) the combined effects of loss, noise, and talker echo. In addition, models were derived for the relationship between transmission rating and subjective opinion.

The procedures used in the analysis of the subjective-test results and the derivation of the transmission-rating scale are described in Sections V and VI. Although the procedures are somewhat complex for manual calculation, they are easily handled on a digital computer and have been found to provide a convenient and useful representation for a large variety of test data.

Mathematical expressions for the models are summarized in Table II. The derivations of these expressions are also given in Sections V

---

* The several subjective tests, results of which were used in deriving the model, were conducted with different circuit-noise values and with telephone sets operating at different EARS receiving efficiencies. To enable combination of results from the different tests to be made, it was necessary to express all circuit-noise values in terms of a telephone set with reference receiving sensitivity. The value of 26 dB was chosen because it is approximately the receiving efficiency of a customer loop consisting of a Western Electric 500-type telephone set,[7] a short-line facility, and a standard central office feeding bridge. Noise values are given in dBrnC.[9,10]

and VI. The remainder of the present section provides a general description and graphical presentation of the models.

2.1 Connection loudness loss and circuit-noise model

Transmission rating as a function of connection loudness loss and message-circuit noise is shown in Fig. 1. The curves were plotted using

Table II—Models for estimating subjective reaction to loss, noise, and echo

The models, in terms of a transmission-rating scale, for loss and noise (R_{LN}), echo (R_E), and loss, noise, and echo (R_{LNE}) are:

$$R_{LN} = 147.76 - 2.257\sqrt{(L_e - 7.2)^2 + 1} - 2.009N_F + 0.02037(L_e)N_F \qquad (1)$$

$$R_E = 95.01 - 53.45\log_{10}\{(1 + D)/\sqrt{1 + (D/480)^2}\} + 2.277E \qquad (2)$$

$$R_{LNE} = \frac{R_{LN} + R_E}{2} - \sqrt{\left[\frac{R_{LN} - R_E}{2}\right]^2 + (10)^2}, \qquad (3)$$

where L_e = Acoustic-to-acoustic loudness loss (in dB) of an overall telephone connection, determined using the Electro-Acoustic Rating System (EARS) method,
N = Circuit noise (in dBrnC) at the input to a set with a receiving-loudness rating of 26 dB, determined using the EARS method,
N_F = Total noise in dBrnC resulting from power addition of the circuit noise, N, and 27.37, both in dBrnC,
D = Round-trip echo-path delay (in milliseconds), and
E = Acoustic-to-acoustic loudness loss (in dB) of the echo path, determined using the EARS method.

The proportion of comments good or better (GoB) or poor or worse (PoW) are computed from R by:

$$GoB = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{A} e^{-t^2/2}dt$$

$$PoW = \frac{1}{\sqrt{2\pi}}\int_{B}^{\infty} e^{-t^2/2}dt,$$

where A and B are given in the table below for the various data bases.

Data Base	A	B
MH	$(R - 64.07)/17.57$	$(R - 51.87)/17.57$
HO1	$(R - 77.44)/17.07$	$(R - 60.70)/17.07$
HO2	$(R - 73.74)/15.68$	$(R - 58.03)/15.68$
Echo1	$(R - 75.05)/14.30$	$(R - 58.95)/14.30$
Echo2	$(R - 66.66)/11.84$	$(R - 53.33)/11.84$

The parameters A and B have been derived from opinion distributions, in terms of fit mean, μ, and fit standard deviation, σ, and then expressed as a function of R. Alternatively, the models can be expressed as follows:

$$\mu_{\mathrm{MH}} = (R - 21.37)/12.20 \qquad \sigma_a = 1.44$$

$$\mu_{\mathrm{HO1}} = (\mu_{\mathrm{MH}} + 0.206)/1.372 = (R - 18.86)/16.74 \qquad \sigma_a = 1.02$$

$$\mu_{\mathrm{HO2}} = (\mu_{\mathrm{MH}} + 0.215)/1.288 = (R - 18.75)/15.71 \qquad \sigma_a = 0.998$$

$$\mu_{\mathrm{E1}} = (R - 18.7)/16.1 \qquad \sigma_a = 0.888$$

$$\mu_{\mathrm{E2}} = (R - 20)/13.33 \qquad \sigma_a = 0.888$$

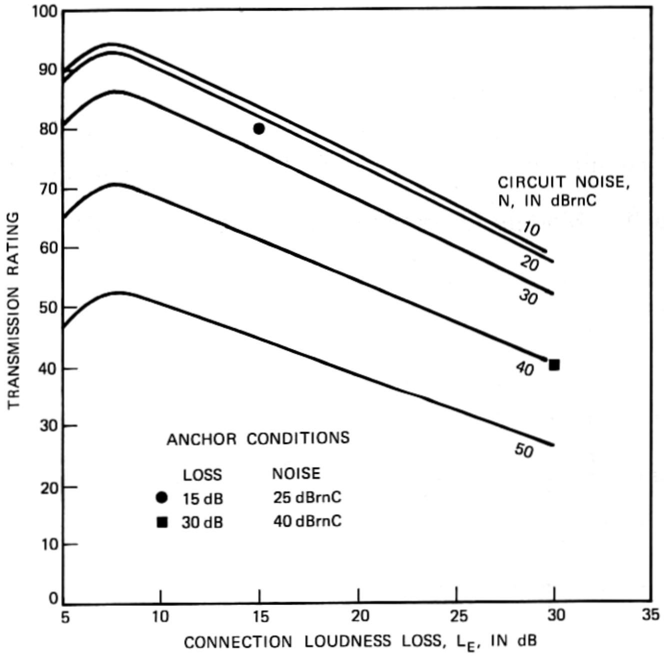

Fig. 1—Transmission rating for loss and noise.

eq. (1) below, which is also given in Table II.

$$R_{LN} = 147.76 - 2.257\sqrt{(L_e - 7.2)^2 + 1} - 2.009N_F + 0.02037(L_e)N_F, \quad (1)$$

where

L_e = Acoustic-to-acoustic loudness loss (in dB) of an overall telephone connection

N = Circuit noise (in dBrnC) at the input to a set with a receiving loudness rating of 26 dB

N_F = Power addition of the circuit noise, N, and 27.37 dBrnC.

The transmission-rating scale was derived, so that it is anchored at two points, as shown in Table III. These anchor points were selected to be well separated in quality, but within the range of conditions that are likely to be included in a test. Transmission ratings for other combinations of connection loudness loss and circuit noise are relative to those for the two anchor points. The rating values are such that most telephone connections will have positive ratings between 40 and 100, with the higher rating denoting higher quality. For most engineering applications sufficient accuracy can be achieved by the use of

Table III — Anchor conditions for the transmission-rating scale

Connection Loudness Loss (dB)	Circuit Noise (dBrnC)	Transmission Rating
15	25	80
30	40	40

whole numbers on the transmission-rating scale. For example, in the 1965 loss and noise tests, conditions with a transmission rating of approximately 80 were considered good or excellent by 80 percent of the subjects, while a transmission rating of 40 was considered good or excellent by only 10 percent.

2.2 Talker-echo model

Transmission rating as a function of talker-echo path loss and delay is shown in Fig. 2. The curves were plotted using eq. (2) below, which is also given in Table II. This equation was derived to exclude the effects of circuit noise and connection loudness loss.

$$R_E = 95.01 - 53.45 \log_{10} \{[1 + D]/\sqrt{1 + (D/480)^2}\} + 2.277E, \quad (2)$$

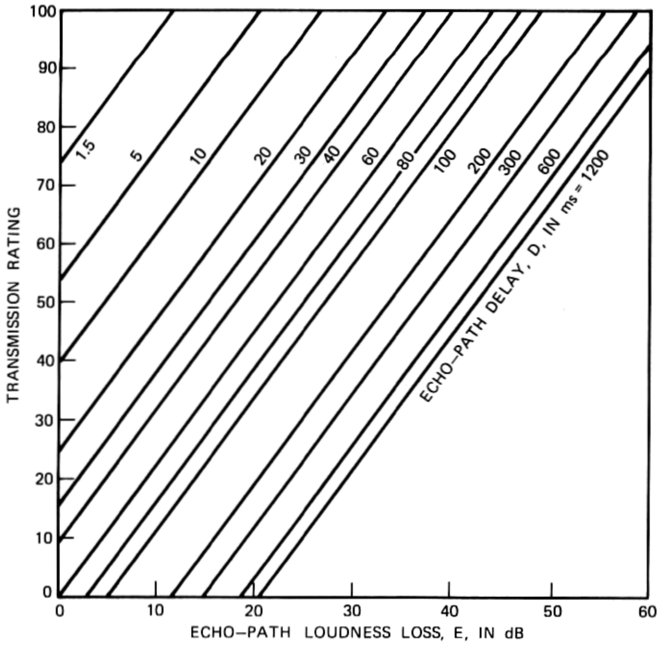

Fig. 2—Transmission rating for talker echo.

where

D = echo-path delay (in ms) and
E = acoustic-to-acoustic loudness loss (in dB) of the echo path.

The curves of Fig. 2 demonstrate the dependence of transmission quality on the two talker-echo path parameters, loss and delay, for connections where talker echo is important.

2.3 Connection loudness loss, circuit noise, and talker-echo model

Transmission ratings for the combined effects of connection loudness loss, circuit noise, echo-path loudness loss and echo-path delay are obtained from eq. (3) below, which is also given in Table II.

$$R_{LNE} = \frac{R_{LN} + R_E}{2} - \sqrt{\left(\frac{R_{LN} - R_E}{2}\right)^2 + (10)^2}, \qquad (3)$$

where

R_{LNE} = transmission rating for the combined effects of connection loudness loss, circuit noise, and talker echo,
R_{LN} = transmission rating for connection loudness loss and circuit noise
R_E = transmission rating for echo-path loudness loss and delay.

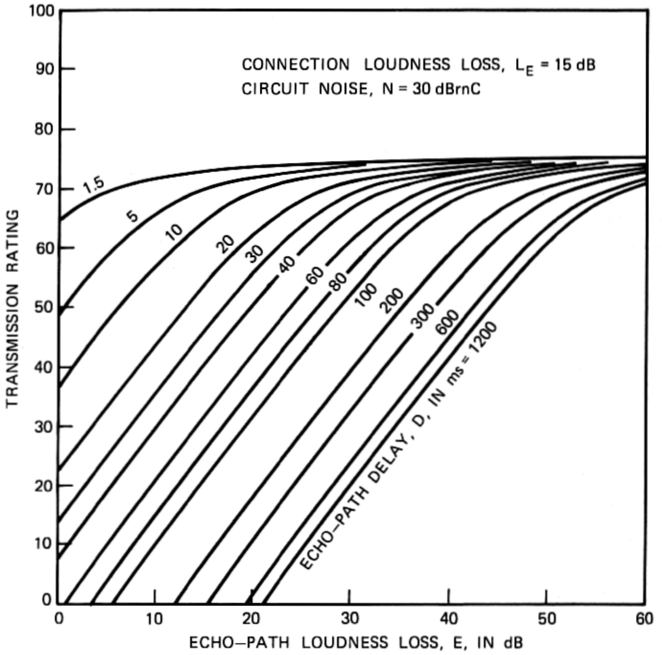

Fig. 3—Transmission rating for loss, noise, and talker echo.

Figure 3 illustrates curves generated by means of the above relationship for the transmission rating as a function of echo-path loudness loss and delay in a connection with a connection loudness loss of 15 dB and circuit noise of 30 dBrnC. For other values of connection loudness loss and circuit noise, the curves would become asymptotic to higher or lower values of *R* in accordance with the curves of Fig. 1.

2.4 Subjective-opinion models

Subjective opinion in terms of the proportion of ratings in each of the five categories (E, G, F, P, U) for a condition having a given transmission rating has been found to depend on various factors, such as the subject group, the range of conditions presented in a test, the year in which the test was conducted, and whether the test was conducted on conversations in a laboratory environment or on normal telephone calls. For the major tests on which the transmission-rating model is based, the observed relationship between subjective judgments and transmission rating can be represented as shown in Figs. 4 and 5 which are plotted from the equations for *GoB* and *PoW* of Table II.

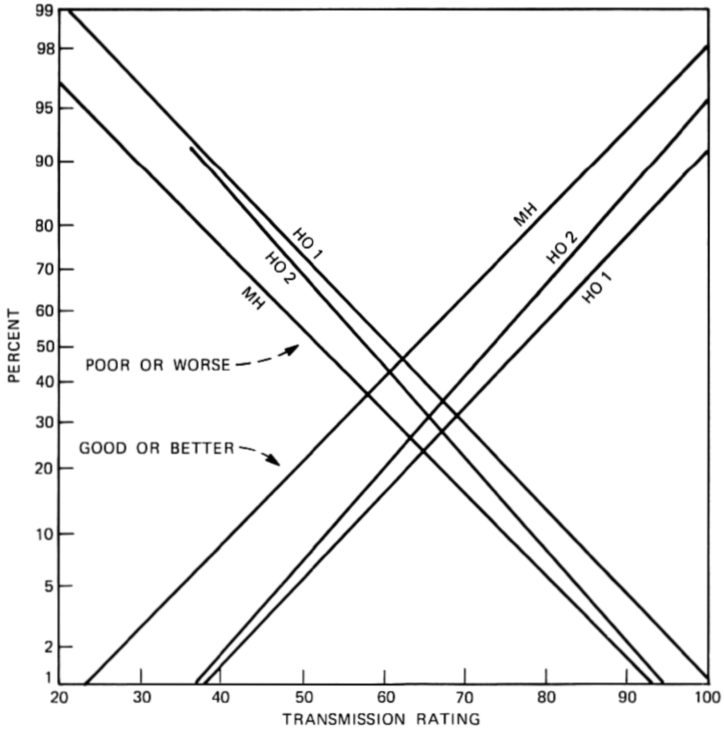

Fig. 4—Subjective opinion as a function of transmission rating for the connection loudness loss and circuit-noise tests.

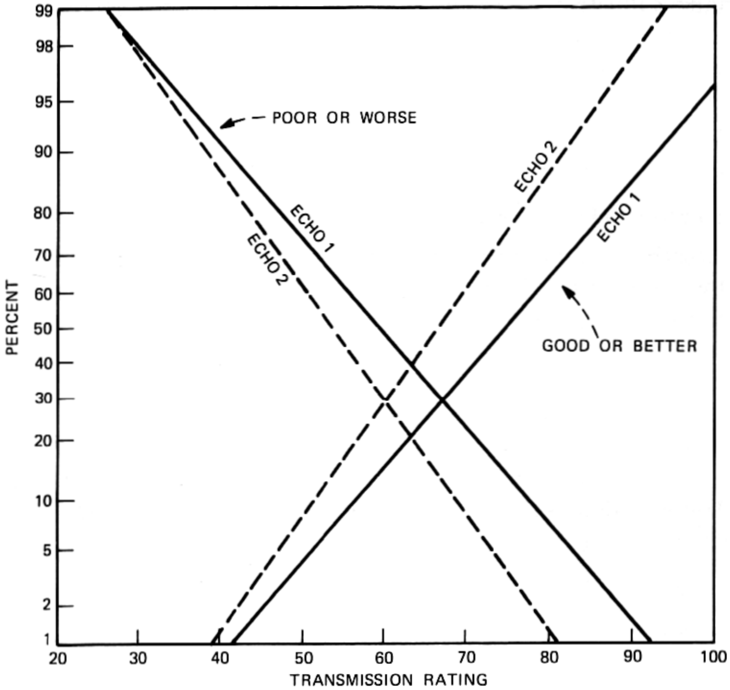

Fig. 5—Subjective opinion as a function of transmission rating for the echo tests.

Contours of constant percent good or better and percent poor or worse are given in Figs. 6 and 7, respectively, for the loss-noise results. These contours were computed from eq. (1) and the equations for *GoB* and *PoW* of Table II using the MH data base. Similar contours could be generated based on one of the other tests. However, the Murray Hill base is being used for current network planning studies for consistency with earlier studies. In addition, the opinion results from this test appear to be in close agreement with data obtained from customer interviews on typical DDD toll connections. Thus, at the time of publication, the Murray Hill base is recommended for conversion of the transmission ratings to subjective ratings. Eventually, we hope to determine other expressions for *A* and *B* (see Table II) that will provide even better agreement with customer interviews on various types of telephone connections.

2.5 Use of the models

The models summarized in preceding sections can be used to estimate transmission quality for telephone connections. The examples given below are based on representative 500-type telephone sets con-

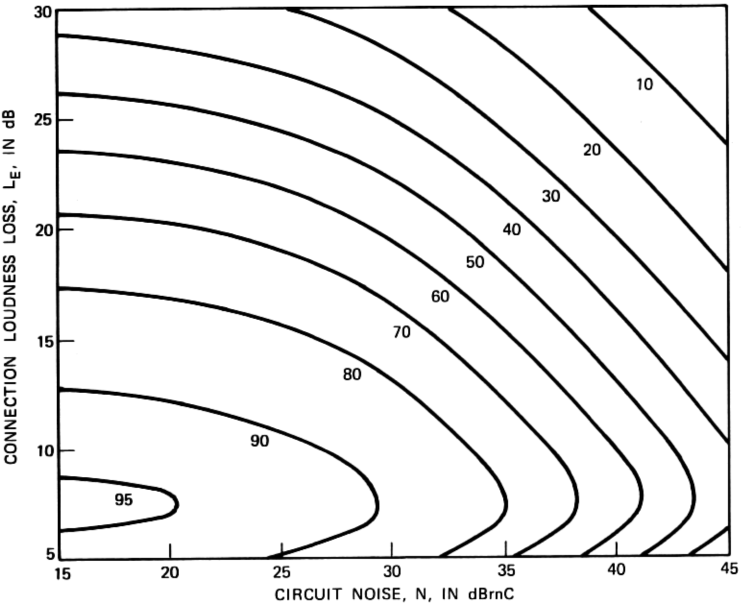

Fig. 6—Subjective-opinion contours of percent good or better at the MH base for loss and noise.

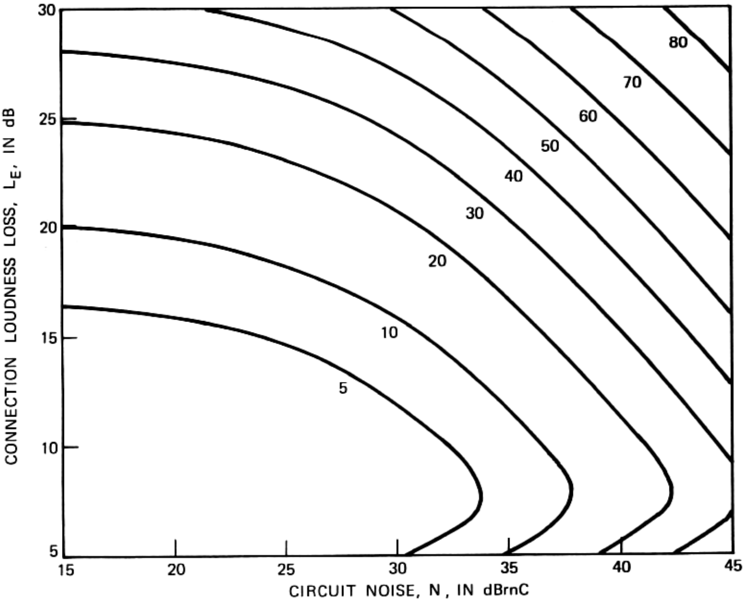

Fig. 7—Subjective-opinion contours of percent poor or worse at the MH base for loss and noise.

nected by specified lengths of 26-gauge nonloaded cable to the local Class 5 office.[1] The loss for such an arrangement is expressed in terms of:

TLR = Transmitting loop rating (in dB), which describes the loudness conversion efficiency in terms of an acoustic signal applied at the telephone set and the resulting electric voltage at the local Class 5 office.

RLR = Receiving loop rating (in dB), which describes the loudness-conversion efficiency in terms of an electric circuit voltage applied at the local Class 5 office and the resulting acoustic pressure received at the telephone set.[8]

The connection loudness loss, L_e, for a local call is the sum of TLR and RLR. For calls between customers served by different local Class 5 offices, L_e is the sum (TLR + RLR) plus the 1000-Hz loss between the offices.

The circuit noise, N (in dBrnC), as used in the model is in terms of a reference telephone set that has an RLR of 26 dB. Noise levels, as typically expressed at the line terminals of a telephone set, need to be corrected to the reference set. For the 500-type telephone set, the conversion factor is about 4 dB (increase in noise) and is nearly independent of loop length.

The effect of talker echo depends on the characteristics of the echo path. Generally, the dominating path is that from the talking customer to the distant Class 5 office and return, and is referred to as the far-end echo path.[11] The round-trip delay of this path, D, is taken to be the time required for a speech signal to go from the originating Class 5 office to the distant Class 5 office and return. (Delay in the customer's telephone set and loop is neglected as, with present plant, it is usually insignificant.) The loudness loss of the echo path, E, is the sum of TLR, RLR, and the echo-path loss from the originating Class 5 office to the distant Class 5 office and return.

Considerations of the preceding paragraphs provide the basis for several examples demonstrating use of the models. Details concerning computation for these examples are given in Appendix A. Results of these examples are summarized in Table IV. The examples are simplified representations of connections devised to illustrate application of the models and, thus, the results only approximately describe the performance of actual connections. (Methods of obtaining more accurate connection representations are covered in Ref. 1.)

Comparison of the results for Examples 1, 2, and 3 show that about optimum performance for local connections occurs for medium loops. The performance is below optimum for short and long loops, the former because the loss is lower, the latter because the loss is higher (see Figs. 1, 6, and 7).

Table IV—Examples of performance estimates obtained using the models

Example	L_e (dB)	N (dBrnC)	D (ms)	E (dB)	R_{LN}	R_e	R_{LNE}	GoB (%)	PoW (%)
1. Local connection, short loops	2.7	27	—	—	78.3	—	—	79.2	6.6
2. Local connection, medium loops	8.4	27	—	—	88.7	—	—	92	1.8
3. Local connection, long loops	16.9	27	—	—	75.5	—	—	74.2	8.9
4. Long-toll connection, medium loops, no talker echo	16.1	32.5	—	—	71	—	—	65.2	13.9
5. Long-toll connection, medium loops, with talker echo	16.1	32.5	37.3	31.7	71	82.6	65.2	52.6	22.4

Comparison of Examples 3 and 4 illustrates the effect of higher loss and noise typically encountered on toll connections.[12] Finally, comparison of Examples 4 and 5 indicates the effect of talker echo, demonstrating the need for echo control. (See Ref. 1 for detailed discussion of loss, noise, and talker echo for toll connections.)

III. DESCRIPTION OF LOSS-NOISE SUBJECTIVE TESTS

Three tests have been conducted using a special test facility called SIBYL to determine subjective reaction to loss and noise on telephone connections. This facility allowed control of transmission parameters during normal business calls of cooperating Bell Laboratories employees.[13–15]

SIBYL was first used at the Murray Hill Bell Laboratories location, and was moved to the Holmdel location in 1966.

The test subjects for the SIBYL studies were Bell Laboratories employees. Prior to the beginning of each of the tests, a list of employees reflecting a preselected makeup of age, sex, etc., was obtained, and the employees contacted to solicit their participation in the test. Upon obtaining agreement, their telephone lines were routed through SIBYL which could handle up to 100 subject lines.

An overall connection with SIBYL inserted is shown on Fig. 8. At the left is a subject's (participating employee's) telephone set. In close physical proximity (less than about 1500 feet of two-conductor cable) is SIBYL which converts the two-wire transmission path into a four-wire transmission path and separates signals transmitted from the subject's telephone set and signals received at the same telephone set. Separation of the signal paths in this manner permits (*i*) inserting different impairment values for the two directions of transmission and (*ii*) independent measurement of signals transmitted from and received at the subject's telephone set.

Proceeding from left to right in Fig. 8, the four-wire path is reconverted to a two-wire path in SIBYL, and connected to the serving

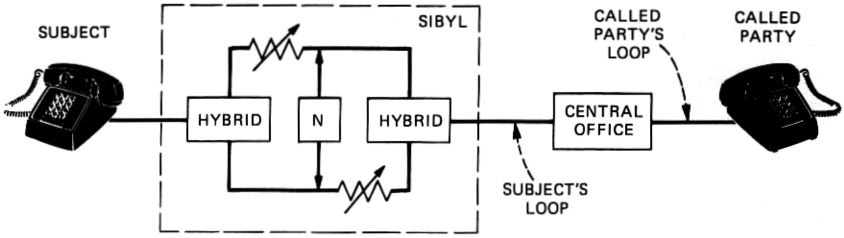

Fig. 8—Diagram of a telephone connection for the SIBYL tests.

central office over a two-wire cable pair which is about 1 mile in length for Murray Hill and about 3 miles in length for Holmdel. The central office switching machine connects the subject's line to the called line dialed by the subject. The called line (approximate lengths as given above for Murray Hill and Holmdel) is terminated with the telephone set of the called party, another Bell Laboratories employee at the same location who is usually not a subject in the experiment.

SIYBL recognized calls internal to the location, and only such calls were included in the experiments reported here. The major reason for this was to retain as much control as possible of the transmission parameters on test calls. Employees' telephone lines at any one location generally were of about the same physical makeup. Thus, variations in transmission parameter values between lines were small, and the values were considered to be identical for all lines. SIBYL then altered the normal parameter values to achieve the parameter values of interest in an experiment.

Restricting the calls to within the location had the further advantage that the subjects physical environment was reasonably uniform, consisting largely of offices with one to four desks and of electronic laboratories. Thus, room noise levels, which could affect the transmission quality of telephone calls, did not vary appreciably.

The subjects were provided with instructions at the beginning of the test, and thereafter, the procedure during a call was generally as shown in Fig. 9. (The procedure of Fig. 9 applies specifically to the Murray Hill test.) The subject initiated the procedure by lifting his handset from the telephone set cradle and dialing a number. If the call was *not* a test call, which was the case for about 85 percent of the calls from the subject group, the call was completed normally. That is, (*i*) the subject received a signal indicating the called line was busy and returned his handset to the cradle, (*ii*) the called party did not answer and the subject returned his handset to the cradle, or (*iii*) the called party answered and they conducted their conversation, after which the subject and called party returned their respective handsets to the cradles.

If the call was a test call and the called line was not busy, a test condition (e.g., a predetermined combination of loss and noise values) was inserted at the beginning of the call, and the conversation proceeded. If, during the call, the test condition was unacceptable to the subject (e.g., the noise was too loud) the subject could dial a digit (e.g., a five) on his telephone set dial. This signaled SIBYL to remove the test condition, after which the call completed normally. (This occurred on only about 1 to 2 percent of all test calls.)

In the more usual case, the subject was either not consciously aware of the test condition or did not find it unacceptable and completed the conversation, then returned the handset to the cradle. A short time thereafter, typically seconds, the subject received ringback—the telephone set ringer emitted a burst of sound—which alerted the subject to rate the transmission quality of the call just completed.

3.1 1965 Murray Hill SIBYL test

In 1965, a test to determine subjective reaction to loss and noise was conducted using the SIBYL facility, which was then located at Bell Laboratories, Murray Hill, New Jersey. The configuration of a typical connection incorporating SIBYL is shown on Fig. 8 as discussed earlier.

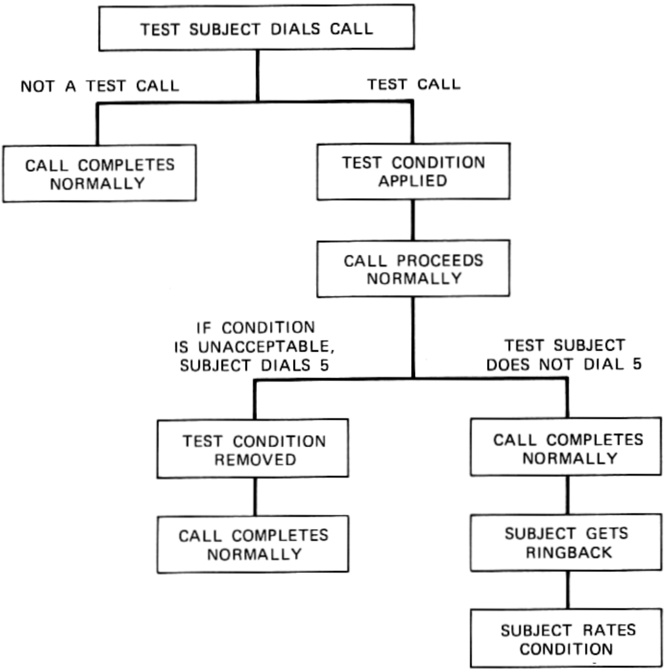

Fig. 9—Procedure for a SIBYL call.

Table V—1965 Murray Hill SIBYL test of loss and noise: number of ratings for each test condition

Connection Loudness Loss (dB)	Circuit Noise (dBrnC)			
	21	28	36	44
5	35	12	24	20
10	61	36	34	34
15	33	33	27	28
20	29	43	23	35
25	35	32	29	24
30	12	12	23	11

The 24 combinations of loss and noise values that were tested are given in Table V together with the number of subjective ratings obtained for each combination. The loss values of Table V represent acoustic-to-acoustic loudness loss (in dB) of the connections, and are numerically equivalent to the electrical losses in terms of which results were published earlier for this test.[6] Circuit noise levels (in dBrnC) were originally reported in terms of (*i*) added noise, and did not include allowance for noise normally present on the connections, and (*ii*) the average receiving sensitivity of Western Electric 500-type telephone sets at the Murray Hill location. The noise levels of Table V represent total noise from all sources and are expressed at the line terminals of a telephone set with reference receiving sensitivity. The condition with loudness loss = 5 dB and noise = 21 dBrnC represents average transmission normally experienced by employees on calls within the Murray Hill location.

The lines connecting employees' telephone sets to the central office were all of about the same length. Thus, the loudness loss of the telephone side tone path,[16,17] strongly dependent on line length, was expected to be about the same for all employees. Side tone path loudness loss was measured on a sample of lines and telephone sets, and found to be about 13 dB. Circuitry was incorporated into SIBYL to ensure that the side tone path loudness loss would also be about 13 dB for subjects' telephone sets when connected to SIBYL.

Room noise was measured at a sample of subject locations using a sound-level meter with A-weighting.[18] The average value was found to be 45 dB(A) with a range of ±2 dB.

Sixty-six employees (subjects) participated in the test. The predetermined combinations of loss and noise were randomly introduced into about 15 percent of all within-location, normal-business calls placed by the subject group during the 3-month test interval. A subject had no prior information that any particular call was being placed over a test connection. The call procedure summarized in Fig. 9 was followed for each call. At the end of each test call, a subject judged overall transmission quality by dialing a 9 for *excellent*, 8 for *good*, 7 for *fair*, and 6 for *poor*. Connections rejected during the call were considered to be *unsatisfactory*.

Test results are given in Table VI for each test condition in terms of the percent of subjects' votes in each of the five rating categories. These results are used in Sections V and VI in deriving the subjective-opinion models.

3.2 1969 Holmdel 1 SIBYL test

As discussed earlier, a second test to determine subjective reaction to loss and noise was conducted in 1969. This test utilized the SIBYL facility which was moved to the Holmdel location of Bell Laboratories in 1966. A major reason for these newer tests was to examine the effect of asymmetric transmission conditions on subjects' ratings. However, the test included 15 symmetric conditions (out of a total of 47 test conditions) and these are considered in this paper.

Table VI — 1965 Murray Hill SIBYL test of loss and noise: test results

Circuit Noise (dBrnC)	Connection Loudness Loss (dB)	Percent of Subjects' Votes in Each Category				
		Excel.	Good	Fair	Poor	Unsat.
21	5	60.0	11.4	14.3	14.3	0
	10	63.9	26.2	3.3	6.6	0
	15	66.7	21.2	12.1	0	0
	20	41.3	27.6	24.1	3.5	3.5
	25	34.3	14.3	31.4	8.6	11.4
	30	8.3	25	33.3	16.7	16.7
28	5	50.0	33.3	16.7	0	0
	10	66.7	30.5	0	2.8	0
	15	54.5	30.3	9.1	6.1	0
	20	25.6	30.2	18.6	16.3	9.3
	25	25.0	15.6	37.5	12.5	9.4
	30	8.3	8.3	33.3	16.8	33.3
36	5	54.2	20.8	16.7	8.3	0
	10	64.7	23.6	5.9	2.9	2.9
	15	33.3	29.6	22.2	14.9	0
	20	21.7	8.7	34.8	21.7	13.1
	25	13.8	3.5	17.2	34.5	31.0
	30	0	13.0	26.1	34.8	26.1
44	5	15.0	30.0	25.0	20.0	10.0
	10	17.6	26.5	11.8	20.6	23.5
	15	17.8	25.0	28.6	14.3	14.3
	20	8.6	5.7	20.0	40.0	25.7
	25	0	4.2	25.0	50.0	20.8
	30	0	0	0	36.4	63.6

Table VII—1969 Holmdel 1 SIBYL test of loss and noise: number of ratings for each test condition

Connection Loudness Loss (dB)	Circuit Noise (dBrnC)			
	22	35	40	45
10	335	48	41	38
15	93	—	—	—
20	65	57	77	77
25	33	34	47	—
30	76	61	—	81

The symmetric combinations of loss and noise values that were tested are given in Table VII together with the number of subjective ratings obtained for each condition. The condition with connection loudness loss = 10 dB and circuit noise = 25 dBrnC was the reference condition that was repeated frequently during the test.

Side tone path loudness loss was measured on a sample of lines and telephone sets, and found to average about 12 dB with a range of ±2 dB both for normal connections and SIBYL connections. Room noise, measured at 12 subject locations, averaged about 42 dB(A) with a range of about ±2 dB.

Seventy-eight employees (subjects) participated in this test. Subjects followed the same procedure during the $2\frac{1}{2}$ month test interval as already described for the Murray Hill test. Comparison of the Holmdel 1 test results of Table VIII with the Murray Hill test results of Table VI shows that in the former, subjects gave lower ratings for approximately equivalent combinations of loss and noise than is the case for the latter. The major differences between the two tests were (*i*) the subject groups, (*ii*) the time difference of about 4 years, and (*iii*) the location. While it is not clear how these differences contributed to differences in results, the comparison suggests that subjects' expectations were higher in the later Holmdel tests.

3.3 1972 Holmdel 2 SIBYL test

A third test to determine subjective reaction to loss and noise was conducted in 1972. This test also utilized the SIBYL facility at the Holmdel location of Bell Laboratories.

A major purpose of the test was to determine whether or not the more critical subjective evaluations found in the Holmdel 1 test would continue to hold. The test included combinations of crosstalk, loss, and noise as well as combinations of loss and noise. Only the latter are considered here because only the results for these conditions are used in the subjective-opinion model for loss and noise.

Table VIII — 1969 Holmdel 1 SIBYL test of loss and noise: test results

Circuit Noise (dBrnC)	Connection Loudness Loss (dB)	Percent of Subjects' Votes in Each Category				
		Excel.	Good	Fair	Poor	Unsat.
22	10	37.9	41.2	13.7	4.5	2.7
	15	29.0	40.9	21.5	5.4	3.2
	20	15.4	26.1	43.1	7.7	7.7
	25	3.0	15.1	27.3	27.3	27.3
	30	2.6	13.2	34.2	30.3	19.7
35	10	12.5	20.8	39.6	25.0	2.1
	20	1.8	10.5	45.6	26.3	15.8
	25	0	2.9	20.6	53.0	23.5
	30	1.6	3.3	16.4	47.5	31.2
40	10	4.9	34.1	24.4	26.8	9.8
	20	0	1.3	14.3	53.2	31.2
	25	2.1	0	12.8	51.1	34.0
45	10	0	7.9	21.0	47.4	23.7
	20	1.3	0	22.1	39.0	37.6
	30	0	0	1.2	33.3	65.5

The 12 symmetric combinations of loss and noise values tested are given in Table IX together with the number of subjective ratings obtained for each combination. These conditions covered about the same range as for the Holmdel 1 (HO1) SIBYL test, except that a condition of lower connection loudness loss of 5 dB was included in the Holmdel 2 (HO2) SIBYL test to match the range of loss in the Murray Hill (MH) test.

Room-noise levels and side tone path loudness loss values were assumed to be identical with those for the HO1 test.

Seventy-four employees (subjects) participated in the test. (None of these employees had been subjects in the HO1 test.) The subjects followed the same procedure as has already been described for the MH

Table IX — 1972 Holmdel 2 SIBYL test of loss and noise: number of ratings for each test condition

Connection Loudness Loss (dB)	Circuit Noise (dBrnC)		
	25	32	42
5	99	70	53
10	1029	58	97
20	50	64	46
30	62	52	4

Table X—1972 Holmdel 2 SIBYL test of loss and noise: test results

Circuit Noise (dBrnC)	Connection Loudness Loss (dB)	Percent of Subjects' Votes in Each Category				
		Excel.	Good	Fair	Poor	Unsat.
25	5	49.5	35.4	13.1	0	2.0
	10	47.2	40.6	10.2	1.3	0.7
	20	14.0	26.0	40.0	18.0	2.0
	30	3.2	3.2	32.3	35.5	25.8
32	5	27.1	35.7	31.4	5.8	0
	10	19.0	39.6	32.8	6.9	1.7
	20	4.7	6.3	48.3	34.4	6.3
	30	1.9	7.7	15.4	51.9	23.1
42	5	11.3	11.3	35.8	34.0	7.6
	10	17.5	17.5	33.1	24.7	7.2
	20	4.3	10.9	19.6	41.3	23.9
	30	0	0	25.0	0	75.0

and HO1 tests, except that the voting procedure was changed. In the HO2 test, subjects were instructed that at the end of each experimental call, when they received ringback, they should rate the overall transmission quality of the call by dialing 9 for *excellent*, 8 for *good*, 7 for *fair*, 6 for *poor*, and 5 for *unsatisfactory*. In addition, they could reject an unacceptable connection during a call by dialing 4. In the latter case, they still received ringback after the call and were asked to rate the quality according to the 5-point scale. (In the MH and HO1 tests, a 4-point scale was used for post-call rating, and the fifth point, unsatisfactory, was assumed for dialed-out calls.)

Results of the HO2 test are given in Table X. These results are in close agreement with those of the HO1 test. Comparison of the test results for the three loss-noise tests are dealt with further in later sections covering derivation of the subjective-opinion models.

IV. DESCRIPTION OF TALKER-ECHO SUBJECTIVE TESTS

Talker echo occurs on a telephone connection when a portion of the primary speech signal is reflected at an impedance mismatch at some point in the connection, and returned to the talker delayed in time. The returned signal, talker echo, is defined in terms of echo-path delay and echo-path loudness loss. The echo-path delay that occurs because of the finite propagation velocity of the speech signal over transmission facilities and equipments is the time it takes the speech signal to traverse the path from the talker's lips to the point of impedance mismatch, then back to the talker's ear. Echo-path loudness loss represents the amount by which the talker's speech signal is attenuated when traversing the same path.

Four tests were conducted to determine subjective reaction to talker echo. Three of these tests—identified as the 1966 Laboratory Echo Test, the 1968 Laboratory Echo Test, and the 1970 Laboratory Echo Test—were conducted under laboratory conditions where the experimenter could closely control conditions. The laboratories used in these tests included rooms that were acoustically designed to muffle both internal and external noise.[15] The fourth test, identified as the 1970 SIBYL Echo Test, was conducted using the SIBYL facility.

The 1966 and 1968 Laboratory Echo Tests and the 1970 SIBYL Echo Tests were designed to study subjective reaction at short echo delays (<100 ms) such as might be encountered on long terrestrial connections. The 1970 Laboratory Echo Test considered the effects of long delays that might be encountered on connections using one- and two-hop synchronous-orbit satellite connections.

4.1 1966 laboratory echo test

The test conducted in 1966 to determine subjective reaction to talker echo utilized the test system shown in block diagram form on Fig. 10. This system provided (*i*) a fixed sidetone path with a loudness loss of about 9 dB, (*ii*) an echo path by means of which the subject heard his own voice delayed in time and attenuated under control of the experimenter, and (*iii*) a transmission path from the test administrator to the subject that had a loudness loss of 14 dB. (Transmission from the subject to the test administrator was obtained by means of an intercom system.) The administrator and subject were located in separate rooms for which the ambient room noise was about 35 dB(A), presumably sufficiently low so as to not affect subjects' ratings. Circuit noise was held constant at 28 dBrnC.

The subject was first given four practice conditions to illustrate the range of transmission quality. Then the actual test conditions were presented. The test incorporated 30 conditions, five different values of

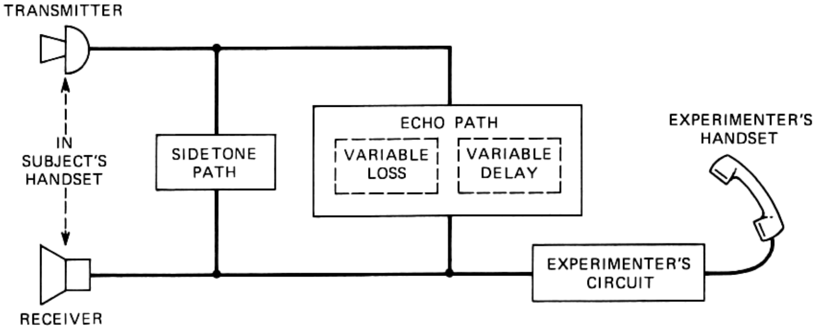

Fig. 10—Test system for the 1966 and 1968 laboratory echo tests.

echo-path delay each with six different values of echo-path loudness loss. These conditions were presented in random order and each subject evaluated each condition once.

At the beginning of each test session, a subject was seated in the test room and given general instructions for the test by the administrator speaking over the system of Fig. 10. The administrator's talking level was held constant. For each of the selected test conditions, the administrator spoke phonetically balanced sentences with the subject repeating each sentence immediately afterward. This continued until the subject arrived at a rating (excellent, good, fair, poor, unsatisfactory) for the test condition. Then the next condition was administered.

The test conditions and the results obtained for the 29-member subject group are given in Table XI. These results show that for any given echo-path delay, the transmission quality improves with increasing echo-path loudness loss. Also, the data indicate that for any given echo-path loudness loss, the transmission quality is degraded with increasing echo-path delay.

4.2 1968 laboratory echo test

This test, conducted in 1968, was designed on the basis of results obtained in the 1966 Laboratory Echo Test. The test system was the same as discussed in the preceding section.

Side tone path loudness loss was constant at 12 dB. Ambient room noise was about 35 dB(A).

The test incorporated 93 conditions. Three of these were base conditions at three different circuit-noise levels. The remaining 90 conditions represented various combinations of circuit-noise level, echo-path delay, and echo-path loudness loss. These conditions were arranged in random order and presented to each test subject in two sessions to avoid subject fatigue.

The procedure for each subject followed that discussed in Section 4.2 except that only three practice conditions were used.

The test conditions and the results obtained for the 30-member subject group are given in Table XII. These test results are reported in part 1 of Annex 4 to Question 6/XII in Ref. 19.

4.3 1970 laboratory echo test

Tests to determine subjective reaction to echo on circuits with echo-path delays of 0 ms, 65 ms, 600 ms, and 1200 ms were conducted in 1970. The tests consisted of a total of 25 conditions, many of which included echo suppressors. One condition at 0 ms, five conditions at 600 ms, and four conditions at 1200 ms did not employ echo sup-

Table XI — 1966 laboratory echo-test results

Echo Path		Percent of Subjects' Votes in Each Category				
Delay (ms)	Loudness Loss (dB)	Excel.	Good	Fair	Poor	Unsat.
20	10	3.5	3.5	41.3	37.9	13.8
	15	0	10.3	48.3	37.9	3.5
	20	6.9	27.6	51.7	13.8	0
	25	41.4	48.3	10.3	0	0
	30	44.8	48.3	6.9	0	0
	35	51.6	44.9	0	3.5	0
36	16.5	3.5	3.5	24.0	55.2	13.8
	21.5	3.5	20.7	20.7	51.6	3.5
	26.5	10.3	58.6	27.6	3.5	0
	31.5	20.7	58.6	17.2	3.5	0
	36.5	41.3	51.7	3.5	3.5	0
	41.5	62.0	31.0	3.5	3.5	0
56	26	3.5	3.5	55.1	34.4	3.5
	31	10.3	17.3	51.7	20.7	0
	36	10.3	34.5	51.7	3.5	0
	41	24.1	48.3	27.6	0	0
	46	51.7	41.3	3.5	0	3.5
	51	55.2	41.3	0	0	3.5
72	29	3.5	6.9	37.9	51.7	0
	34	0	31.0	51.7	17.2	0
	39	0	48.3	48.2	3.5	0
	44	20.7	51.7	27.6	0	0
	49	41.4	37.9	20.7	0	0
	54	44.7	48.3	3.5	3.5	0
90	34	0	10.3	48.3	37.9	3.5
	39	6.9	31.0	48.3	10.3	3.5
	44	10.3	48.3	41.4	0	0
	49	24.1	58.7	10.3	6.9	0
	54	41.4	44.8	13.8	0	0
	59	55.2	41.3	3.5	0	0

pressors. Results for these conditions were used in deriving the echo model.

The test system, shown on Fig. 11, provided for two-way conversation between pairs of subjects. The two ends of the test system were located in separate, acoustically treated rooms for which the ambient room noise was about 38 dB(A). Each subject was able to hear his own voice by means of *(i)* a side tone path with loudness loss of 12 dB and *(ii)* an echo path.

One hundred pairs of subjects participated in the tests; 50 of these evaluated the 600-ms delay case, the other 50 evaluated the 1200-ms delay case. The two members of a pair (they were acquainted) were located in the two separate test rooms. Prior to a test session, they were instructed that they should discuss a subject of mutual interest

Table XII—1968 laboratory echo-test results—30 ratings (subjects) per condition

Circuit Noise (dBrnC)	Echo Path		Percent of Subjects' Votes in Each Category				
	Delay (ms)	Loudness Loss (dB)	Excel.	Good	Fair	Poor	Unsat.
18.0	—	100	63.4	30.0	3.3	0	3.3
18.0	1.5	0.4	3.3	33.3	36.7	23.4	3.3
		5.4	26.7	43.3	23.3	6.7	0
		10.4	33.3	46.7	10.0	10.0	0
		15.4	60.0	26.7	13.3	0	0
		20.4	33.3	46.7	13.3	6.7	0
		25.4	63.3	36.7	0	0	0
18.0	20.0	5.5	0	3.3	3.3	23.4	70.0
		10.5	3.3	0	13.3	40.0	43.4
		15.5	0	10.0	36.7	23.3	30.0
		20.5	33.3	30.0	16.7	13.3	6.7
		25.5	36.7	36.7	23.3	3.3	0
		30.5	26.7	56.7	6.6	10.0	0
18.0	56.0	15.6	0	0	6.7	33.3	60.0
		20.6	0	0	10.0	36.7	53.3
		25.6	3.3	6.7	26.7	36.6	26.7
		30.6	3.3	20.0	40.0	33.4	3.3
		35.6	20.0	43.3	26.7	10.0	0
		40.6	33.3	46.7	16.7	0	3.3
18.0	90.0	24.5	0	0	6.7	40.0	53.3
		29.5	3.3	10.0	20.0	36.7	30.0
		34.5	3.3	13.3	43.3	36.8	3.3
		39.5	23.3	26.7	16.7	30.0	3.3
		44.5	30.0	40.0	23.4	3.3	3.3
		49.5	50.0	40.0	10.0	0	0
28.0	—	100	26.6	60.0	6.7	6.7	0
28.0	1.5	0.4	3.3	26.8	43.3	13.3	13.3
		5.4	10.0	43.3	40.0	6.7	0
		10.4	16.7	56.7	20.0	6.6	0
		15.4	16.7	56.7	16.7	6.6	3.3
		20.4	6.7	56.7	30.0	3.3	3.3
		25.4	23.3	60.0	13.4	0	3.3
28.0	10.0	3.0	0	3.4	23.3	33.3	40.0
		8.0	0	3.3	43.3	36.7	16.7
		13.0	3.3	20.0	33.3	26.7	16.7
		18.0	20.0	56.7	13.3	3.3	6.7
		23.0	3.3	46.7	43.3	6.7	0
		28.0	10.0	60.0	26.7	3.3	0
28.0	20.0	5.5	0	0	13.3	33.3	53.4
		10.5	3.3	6.7	33.3	26.7	30.0
		15.5	10.0	3.3	23.3	43.4	20.0
		20.5	3.3	23.4	60.0	10.0	3.3
		25.5	13.3	56.7	23.3	0	6.7
		30.5	10.0	56.7	23.3	6.7	3.3

Table XII — Continued

Circuit Noise (dBrnC)	Echo Path		Percent of Subjects' Votes in Each Category				
	Delay (ms)	Loudness Loss (dB)	Excel.	Good	Fair	Poor	Unsat.
28.0	36.0	11.0	0	0	10.0	16.7	73.3
		16.0	0	0	6.7	33.3	60.0
		21.0	0	0	26.7	50.0	23.3
		26.0	3.3	20.0	60.0	16.7	0
		31.0	6.7	50.0	33.3	6.7	3.3
		36.0	13.3	60.0	20.0	6.7	0
	56.0	15.6	0	3.3	10.0	20.0	66.7
		20.6	3.3	0	6.7	36.7	53.3
		25.6	0	3.3	40.0	30.0	26.7
		30.6	0	13.3	50.0	33.4	3.3
		35.6	3.3	46.7	36.7	13.3	0
		40.6	16.7	50.0	33.3	0	0
	72.0	21.5	3.3	3.3	6.7	33.3	53.4
		26.5	0	0	13.3	63.4	23.3
		31.5	3.3	3.3	33.4	40.0	20.0
		36.5	6.7	36.6	43.3	6.7	6.7
		41.5	6.7	36.6	46.7	10.0	0
		46.5	23.3	56.7	13.3	6.7	0
	90.0	24.5	0	0	3.3	33.3	63.4
		29.5	3.3	0	26.7	50.0	20.0
		34.5	6.7	23.4	33.3	33.3	3.3
		39.5	6.7	30.0	46.6	16.7	0
		44.5	3.3	46.7	43.3	6.7	0
		49.5	10.0	50.0	36.7	0	3.3
38.0	—	100	0	26.7	53.3	13.3	6.7
38.0	1.5	0.4	3.3	0	26.7	46.7	23.3
		5.4	0	3.3	30.0	53.4	13.3
		10.4	0	26.7	50.0	13.3	10.0
		15.4	3.3	26.7	50.0	20.0	0
		20.4	0	13.3	56.7	20.0	10.0
		25.4	6.7	10.0	40.0	40.0	3.3
	20.0	5.5	0	0	3.3	20.0	76.7
		10.5	0	0	3.3	36.7	60.0
		15.5	0	3.3	20.0	43.4	33.3
		20.5	0	6.7	43.3	43.3	6.7
		25.5	6.7	13.3	46.7	23.3	10.0
		30.5	0	6.7	60.0	20.0	13.3
	56.0	15.6	0	0	3.3	26.7	70.0
		20.6	0	0	3.3	46.7	50.0
		25.6	0	0	20.0	53.3	26.7
		30.6	0	0	53.4	43.3	3.3
		35.6	6.7	16.7	40.0	33.3	3.3
		40.6	0	13.3	46.7	30.0	10.0
	90.0	24.5	0	0	13.3	36.7	50.0
		29.5	0	0	16.7	53.3	30.0
		34.5	0	13.3	36.7	43.3	6.7
		39.5	0	23.3	40.0	36.7	0
		44.5	6.7	26.7	40.0	23.3	3.3
		49.5	3.3	23.3	56.7	10.0	6.7

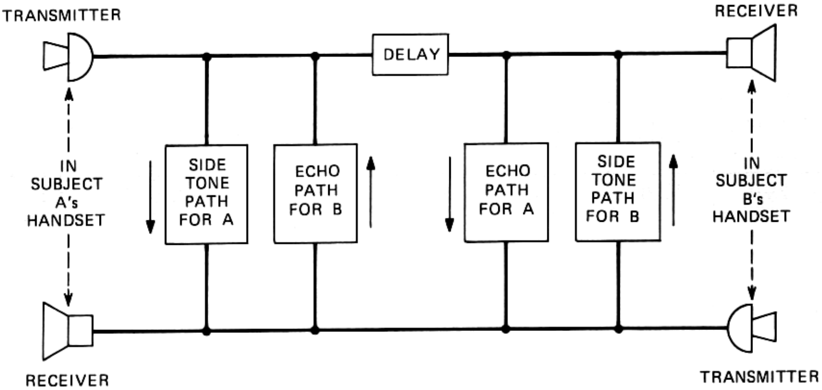

Fig. 11—Test system for the 1970 laboratory echo test.

over the system and told that each test condition would require about 4 minutes of conversation, at the end of which they should separately rate the condition on the 5-point scale: excellent, good, fair, poor, and bad. (These categories were further subdivided into undesignated thirds, resulting in a 15-point scale.)

The test conditions and results are given in Table XIII and are also reported in Part III of Annex 5 to Question 6/XII in Ref. 19. As with tests discussed earlier, these results show that (*i*) at a given echo-path

Table XIII—1970 laboratory echo-test results—Approximately 100 ratings (subjects) per condition

Connection loudness loss = 18 dB
Side tone-path loudness loss = 12 dB
Circuit noise = 33 dBrnC

Echo Path		Percent of Subjects' Votes in Each Category				
Delay (ms)	Loudness Loss (dB)	Excel.	Good	Fair	Poor	Unsat.
—	33.0	9.0	49.0	34.0	8.0	0.0
600	33.0	1.0	1.0	2.0	17.6	78.4
	43.0	2.0	10.8	19.6	39.2	28.4
	53.0	3.9	36.3	44.1	11.8	3.9
	63.0	6.9	47.1	28.4	16.6	1.0
	73.0	6.0	44.0	42.0	6.0	2.0
1200	43.0	1.0	5.0	19.0	38.0	37.0
	53.0	6.0	29.0	40.0	19.0	6.0
	63.0	4.0	45.0	36.0	14.0	1.0
	73.0	7.0	39.0	39.0	13.0	2.0

delay, transmission quality improves with increasing echo-path loudness loss, and (*ii*) at a given echo-path loudness loss, transmission quality is degraded with increasing echo-path delay.

4.4 1970 SIBYL echo test

An echo test was conducted in early 1970 using SIBYL. The purpose of the test was to enable comparison of results obtained from subjects conversing on actual telephone calls to results obtained under laboratory conditions. The results of tests reported in Section 4.2 were used to guide selection of the conditions for the 1970 SIBYL echo test.

Forty-five subjects participated in these tests. Procedures followed by the subjects were the same as those for the loss-noise tests reported in Section 3.3.

Test variables were echo-path delay (three values) and echo-path loudness loss (five values for each delay). In addition, a condition without echo was included as a reference.

Connection loudness loss was 10 dB, the base condition for the Holmdel SIBYL tests reported in Sections 3.2 and 3.3. Circuit noise was 30 dBrnC. Side tone path loudness loss was about 12 dB and average room noise was estimated to be about 42 dB(A).

The test conditions and results are given in Table XIV. As with results of echo tests discussed in preceding sections, these results show that transmission quality is strongly dependent on echo-path delay and echo-path loudness loss.

V. ANALYSIS OF INDIVIDUAL TEST DATA

The raw test results from any individual test provide subjective-opinion information expressible in the form of percent of ratings in each of the five rating categories for each test condition. The utilization of these raw test results in this form for transmission planning is difficult because it is usually necessary to have ratings available for transmission parameter values not specifically included in the tests. Thus, some form of data analysis is frequently applied to obtain graphical or analytical representations of the data that are more convenient for use in transmission planning studies. This can involve simple curve fitting to the raw test data[5,6] or more elaborate models of subject ratings using binomial or other distributions.[20]

For example, Lewinski in Ref. 5 provides separate "smooth" fits for the percentage of responses in the categories excellent, good or better, fair or better, and poor or better as a function of circuit noise. Similarly Sen in Ref. 6 provides mathematical expressions and contours for the percent good or better and percent poor or worse as a function of connection loss and noise. A different approach is suggested by

Table XIV—1970 SIBYL echo test
Connection loudness loss = 10 dB
Circuit Noise = 30 dBrnC

Echo Path		Number of Ratings	Percent of Subjects' Votes in Each Category				
Delay (ms)	Loudness Loss (dB)		Excel.	Good	Fair	Poor	Unsat.
—	∞	183	32.2	36.6	24.6	5.5	1.1
10.0	5.8	51	0	15.7	29.4	47.1	7.8
	10.8	23	0	8.7	30.4	47.8	13.1
	15.8	38	15.8	21.0	39.5	21.0	2.7
	20.8	45	8.9	53.3	20.0	13.3	4.5
	25.8	41	19.5	51.2	29.3	0	0
36.0	14.4	38	0	2.6	18.4	50.0	29.0
	19.4	37	2.7	0	29.7	37.9	29.7
	24.4	51	11.8	35.3	23.5	23.5	5.9
	29.4	31	6.4	35.6	25.8	25.8	6.4
	34.4	46	19.6	36.9	28.3	13.0	2.2
72.0	22.4	35	0	2.9	2.9	65.7	28.5
	27.4	20	10.0	0	15.0	60.0	15.0
	32.4	34	5.9	8.8	29.4	38.2	17.7
	37.4	33	12.1	24.2	30.4	24.2	9.1
	42.4	46	13.1	47.8	30.4	6.5	2.2

Prosser, Allnatt, and Lewis in Ref. 20 where they point out that five separate mathematical functions can be specified, one for each grade on the rating scale, but advocate the desirability of a more convenient and compact representation by means of a single mathematical model that embraces all five functions. They examined various models based on the binomial distributions as well as logistic and gaussian curves. They adopted the second-order binomial as the simplest adequate model to describe the opinion distribution found in their experiment.

We also recognized the advantages of a single mathematical model to represent the distribution of opinion in the five rating categories. The normal density curve was selected as a basis for the model described in the following sections because it provided somewhat greater flexibility in accommodating a variety of standard deviations. Because of the availability of digital computers for the data analysis, the additional computational complexity associated with the normal distribution was not judged to be a problem.

5.1 Analysis method

The subjective-test results for each test condition of connection loudness loss, L_e, and circuit noise, N, form a vote histogram containing the proportion, P_i, of ratings for each of the rating categories, $i = 1, 2, 3, 4, 5$. Rating category 1 represents the unsatisfactory category, 2 represents poor, 3 represents fair, 4 represents good, and 5 represents excellent.

The P_i's for each test condition sum to unity.

The values of P_i for each test condition were used to calculate the mean opinion score (MOS) and sample standard deviation (SIGMOS) as in eqs. (4) and (5), respectively.

$$\text{MOS} = \sum_{i=1}^{5} iP_i \tag{4}$$

$$\text{SIGMOS} = \left[\sum_{i=1}^{5} i^2P_i - (\text{MOS})^2\right]^{\frac{1}{2}}. \tag{5}$$

The vote histogram was represented by a normal density curve with mean, μ, and standard deviation, σ. The area under this curve was divided into five regions, each with area $\hat{P}_i$. The areas were defined as follows: from minus infinity to 1.5 as $\hat{P}_1$, from 1.5 to 2.5 as $\hat{P}_2$, from 2.5 to 3.5 as $\hat{P}_3$, from 3.5 to 4.5 as $\hat{P}_4$, and from 4.5 to infinity as $\hat{P}_5$; the $\hat{P}_i$'s sum to unity. This quantization of the area under the normal curve into five discrete regions was the basis for using the normal curve

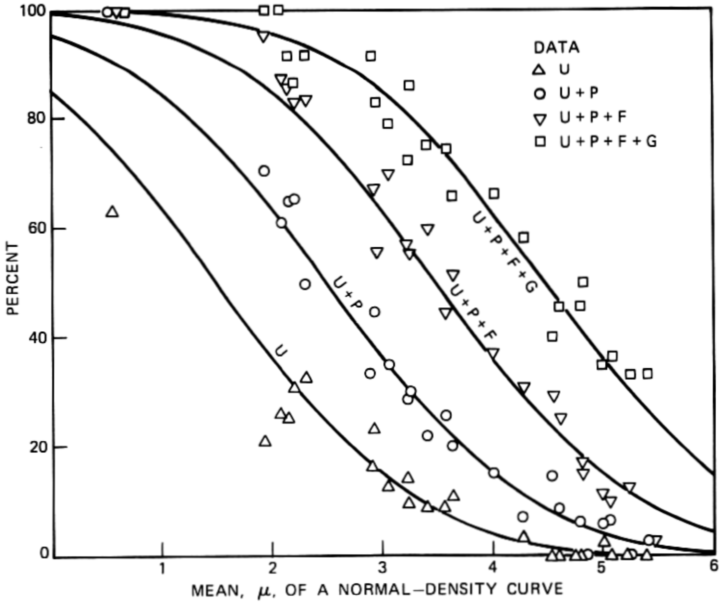

Fig. 12—MH loss-noise test data percentages as a function of the Step 3 normal-density means compared with percentages predicted using the MH standard deviation, 1.44.

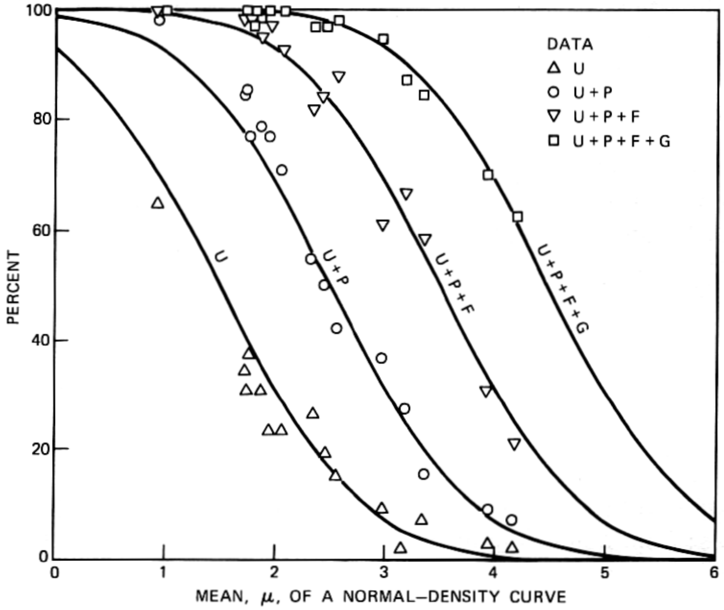

Fig. 13—HO1 loss-noise test data percentages as a function of the Step 3 normal-density means compared with percentages predicted using the HO1 standard deviation, 1.02.

to represent this type of data. These $\hat{P}_i$'s were used to compute the pseudo mean opinion scores (MOS_Q) and sample standard deviations (SIGMOS_Q) in terms of the mean, μ, and standard deviation, σ, of the normal curve as given in eqs. (6) and (7), respectively.

$$\text{MOS}_Q = \sum_{i=1}^{5} i\hat{P}_i = 5 - \sum_{j=1}^{4} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(j+0.5-\mu)/\sigma} \exp\left(-\frac{t^2}{2}\right) dt \right]. \tag{6}$$

$$\begin{aligned} \text{SIGMOS}_Q &= \left[\sum_{i=1}^{5} i^2\hat{P}_i - (\text{MOS}_Q)^2 \right]^{\frac{1}{2}} \\ &= \left\{ 25 - \sum_{j=1}^{4} \left[\frac{2j+1}{\sqrt{2\pi}} \int_{-\infty}^{(j+0.5-\mu)/\sigma} \exp\left(-\frac{t^2}{2}\right) dt \right] \right. \\ &\qquad \left. - (\text{MOS}_Q)^2 \right\}^{\frac{1}{2}}. \end{aligned} \tag{7}$$

Step 1 in the four-step analysis was to find a normal density curve for each test condition such that MOS = MOS_Q and SIGMOS = SIGMOS_Q. These two constraints were used in an iterative computer procedure to determine the values of the mean, μ, and standard deviation, σ, of the normal density curve used to represent the data for each test condition.

Three other criteria were considered in the determination of a normal density curve during the development of this analysis procedure. One criterion was to determine μ and σ such that the proportions good or better ($P_4 + P_5$) and poor or worse ($P_1 + P_2$) were the same for the data and the normal curve. Another was to determine μ and σ to minimize the sum of the squares of the differences between P_1 and $\hat{P}_1$, $P_1 + P_2$ and $\hat{P}_1 + \hat{P}_2$, $P_1 + P_2 + P_3$ and $\hat{P}_1 + \hat{P}_2 + \hat{P}_3$, and finally $P_1 + P_2 + P_3 + P_4$ and $\hat{P}_1 + \hat{P}_2 + \hat{P}_3 + \hat{P}_4$. The third criterion was to determine μ and σ to minimize the sum of the square of the differences $P_i - \hat{P}_i$, $i = 1, 2, \cdots, 5$. The selection of MOS = MOS_Q and SIGMOS = SIGMOS_Q was chosen as the general criterion of fit, since the other criteria were found to be more sensitive to the experimental variability inherent in subjective data of this type, particularly when the number of subjects used is small (less than about 100). This selection was made after applying the several analysis criteria to several hundred sets of test data generated by Monte Carlo simulation.

After all the test conditions were represented by a normal density curve with mean, μ, and standard deviation, σ, an average of the standard deviations for all of the test conditions was computed in Step 2. In determining this average, the individual σ's were weighted in accordance with the number of votes per condition and by the weight-

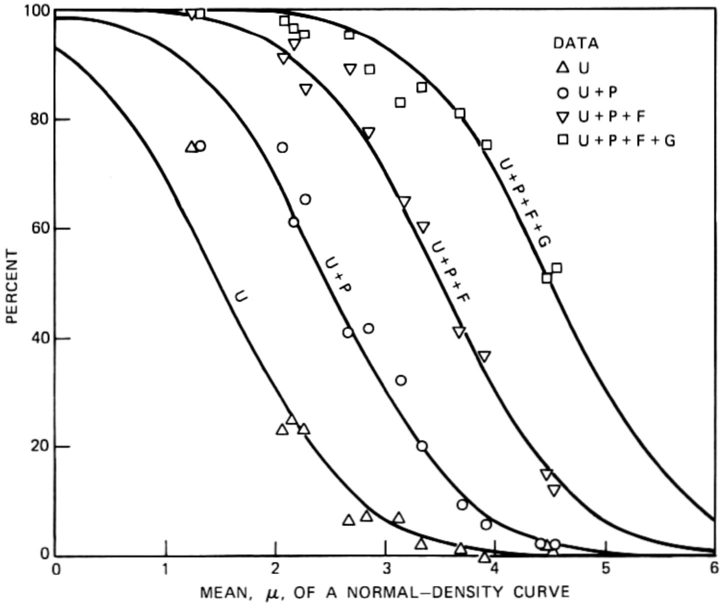

Fig. 14—HO2 loss-noise test percentage as a function of the Step 3 normal-density means compared with percentages predicted using HO2 standard deviation, 0.998.

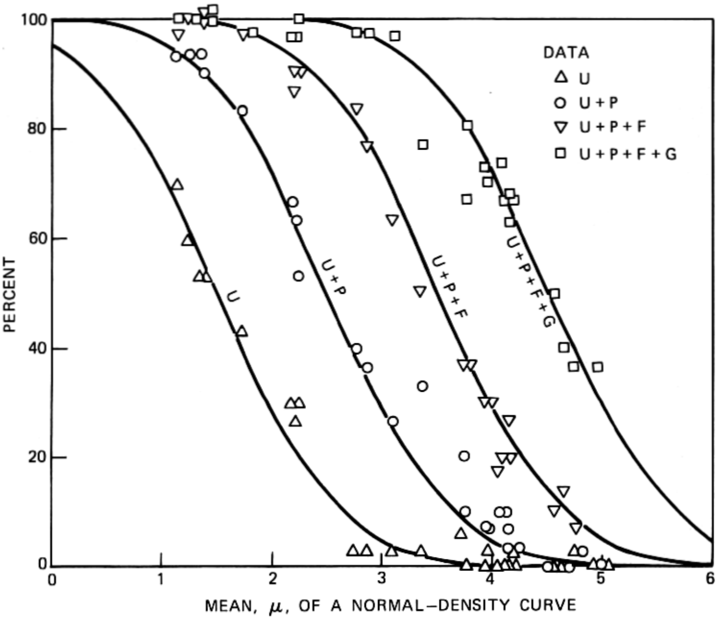

Fig. 15—1968 laboratory echo data percentages (noise = 18 dBrnC) as a function of the Step 3 normal-density means compared with percentages predicted using the echo standard deviation, 0.888.

ing given in eq. (8). This latter weighting was based on the analyses of the Monte Carlo data mentioned previously by examining the variations in the standard deviations as a function of the fit mean, μ.

$$\text{Weighting} = \left[\frac{1}{1 + \frac{(\mu - 3)^2}{3}}\right]^2. \tag{8}$$

The weighted average standard deviation, σ_a, was then used in Step 3 as the standard deviation for all of the normal density curves. Using σ_a, a new mean, μ_a, was computed for each test condition subject to the constraint, MOS = MOS$_Q$. The end result was a family of normal density curves, all with the same standard deviation and with means as determined above.

Figures 12 to 19 illustrate the results from the first three steps in the procedure for several of the tests described in this paper. In these figures the cumulative percent of ratings in four categories—unsatisfactory, unsatisfactory plus poor, unsatisfactory plus poor plus fair, unsatisfactory plus poor plus fair plus good—are plotted against the fit mean, μ_a, determined in Step 3. The solid curves are plotted using the weighted average standard deviation, σ_a, from Step 2. Also shown are the raw data plotted against the respective fit mean, μ_a, for each condition. These figures show that the normal density curves defined by the values, μ_a, and an average standard deviation, σ_a, provide a convenient and simple representation of the raw data for any single test.

In Step 4, the means of all the normal curves are fitted by a suitable analytical function of the test parameters using a least-squares-fit technique.

In summary, the steps involved in the analysis procedure can be described as follows:

Step 1. A normal density curve is used to represent the vote histogram for each test condition such that MOS = MOS_Q and SIGMOS = SIGMOS_Q.

Step 2. The standard deviations of the normal curves of Step 1 for all test conditions under consideration are weighted and averaged to obtain a single value for the standard deviation, σ_a.

Step 3. The single value of standard deviation from Step 2 is used for each test condition as the standard deviation of the

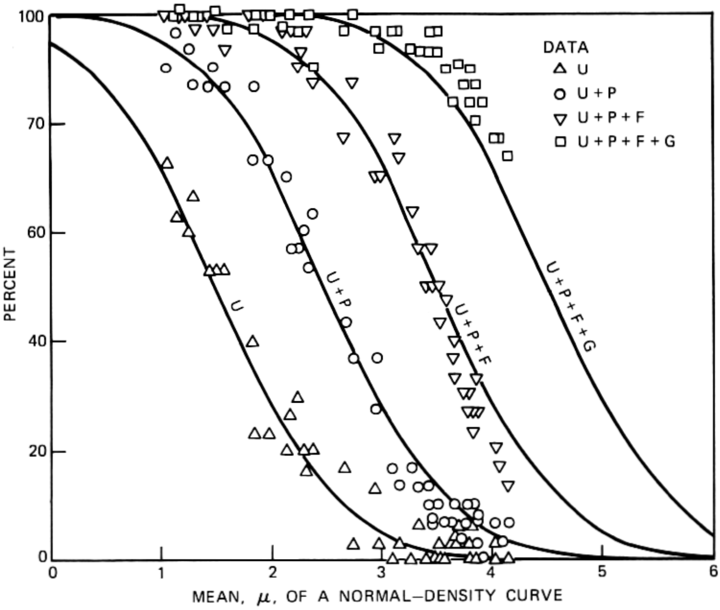

Fig. 16—1968 laboratory echo data percentages (noise=28 dBrnC) as a function of the Step 3 normal-density means compared with percentages predicted using the echo standard deviation, 0.888.

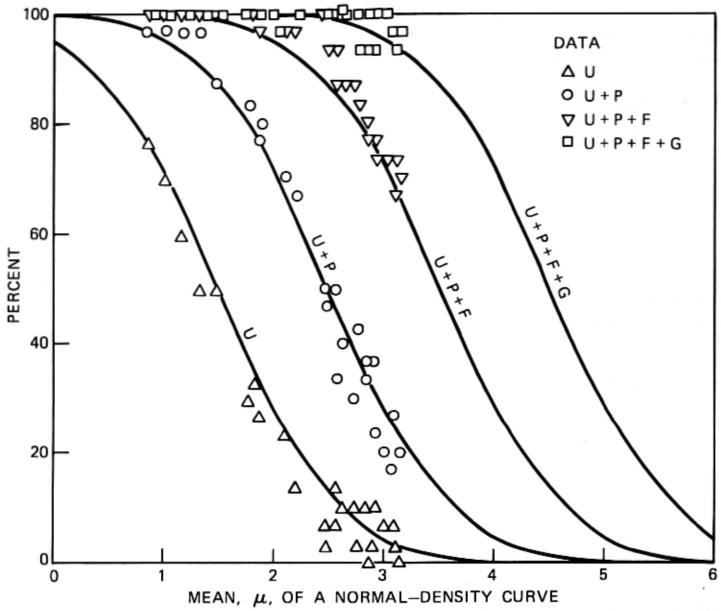

Fig. 17—1968 laboratory echo data percentages (noise=38 dBrnC) as a function of the Step 3 normal-density means compared with percentages predicted using the echo standard deviation, 0.888.

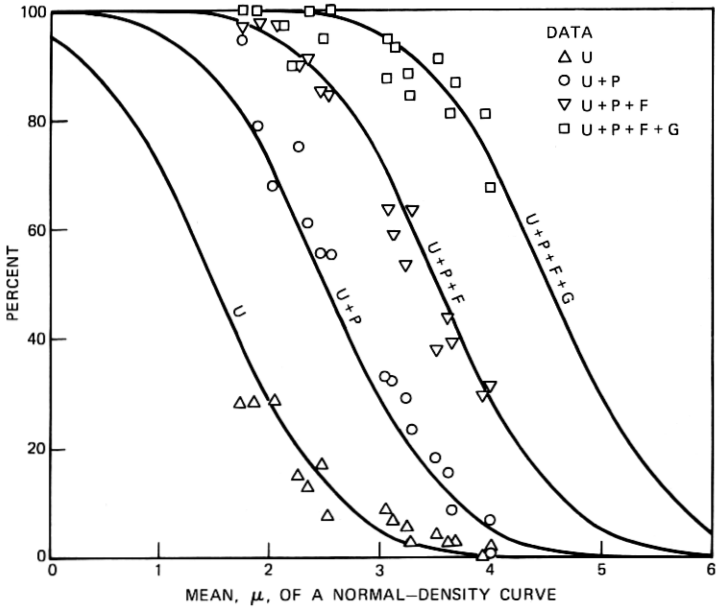

Fig. 18—SIBYL echo data percentages as a function of the Step 3 normal-density means compared with percentages predicted using the echo standard deviation, 0.888.

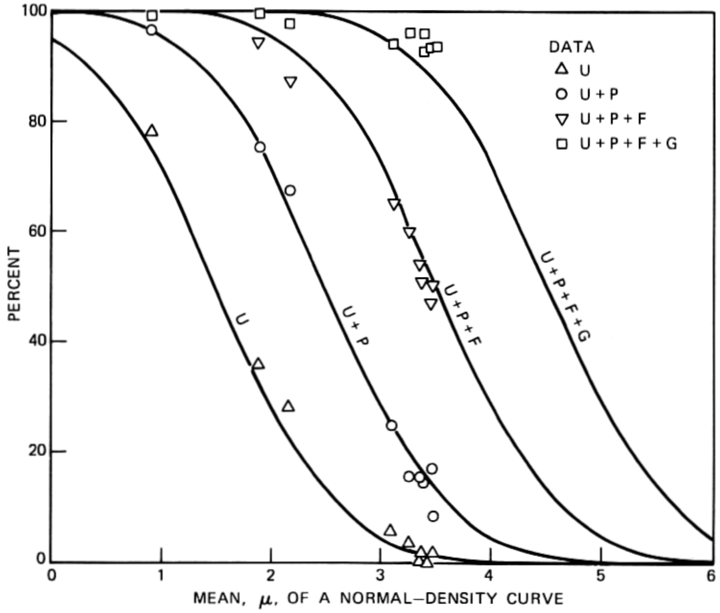

Fig. 19—1970 laboratory echo data percentages as a function of the Step 3 normal-density means compared with percentages predicted using the echo standard deviation, 0.888.

corresponding normal curve, and the mean of the normal curve, μ_a, is adjusted such that MOS = MOS_Q.

Step 4. The means of all normal curves of Step 3 are fitted, on a least-squares-error basis, to an appropriate function of the test parameters.

The results at each step of this analysis are summarized in Table XV for HO2 loss/noise data.

In this table, the results presented for Step 4 are based on the analytical function [eq. (11) in the next section] fitted in Step 4. Comparison of the entries for Step 4 with the entries for the previous steps illustrate the extent to which this analytical function provides a good fit to the test results. For this test, all of the fit means in Step 4 are within approximately 0.2 of the fit means in Step 3. The individual differences have a mean of 0.035 and a standard deviation of 0.13. This agreement is considered reasonable in view of the average standard deviation of 0.998 and about 50 to 70 ratings per condition. Somewhat larger differences were obtained in the MH and HO1 tests where the number of ratings per test condition was smaller.

5.2 Loss-noise analysis

The results obtained by applying the analysis method to the data from the three loss/noise tests are given in eqs. (9), (10), and (11),

Table XV—Summary of results for HO2 loss-noise test at each step in the analysis procedure

Step in the Fit Process	Percentage of Votes					Mean Opinion Score*	Standard Deviation*	Fit Mean	Fit Sigma
	Excel.	Good	Fair	Poor	Unsat.				
Test Condition 1, Loss = 5 dB, Noise = 25 dBrnC									
Raw Data	49.49	35.35	13.13	0.00	2.02	4.303	0.846	—	—
Step 1	51.54	30.94	14.15	3.05	0.32	4.303	0.846	4.543	1.117
Step 2	51.71	33.48	12.77	1.92	0.11	4.348	0.782	4.543	0.998
Step 3	49.16	34.51	13.96	2.23	0.14	4.303	0.801	4.479	0.998
Step 4	50.00	34.18	13.56	2.12	0.13	4.318	0.795	4.500	0.998
Test Condition 2, Loss = 10 dB, Noise = 25 dBrnC									
Raw Data	47.23	40.62	10.20	1.26	0.68	4.325	0.762	—	—
Step 1	48.79	36.59	12.98	1.57	0.06	4.325	0.762	4.472	0.923
Step 2	48.88	34.61	14.10	2.26	0.15	4.298	0.803	4.472	0.998
Step 3	50.40	34.02	13.38	2.07	0.13	4.325	0.792	4.510	0.998
Step 4	50.00	34.18	13.56	2.12	0.13	4.318	0.795	4.500	0.998
Test Condition 3, Loss = 20 dB, Noise = 25 dBrnC									
Raw Data	14.00	26.00	40.00	18.00	2.00	3.320	0.989	—	—
Step 1	11.81	31.46	36.85	16.75	3.14	3.320	0.989	3.333	0.985
Step 2	12.11	31.24	36.45	16.88	3.31	3.320	0.998	3.333	0.998
Step 3	12.11	31.24	36.45	16.88	3.31	3.320	0.998	3.333	0.998
Step 4	12.26	31.38	36.37	16.73	3.26	3.326	0.998	3.340	0.998
Test Condition 4, Loss = 30 dB, Noise = 25 dBrnC									
Raw Data	3.23	3.23	32.26	35.48	25.81	2.226	0.974	—	—
Step 1	1.26	8.78	27.39	36.40	26.16	2.226	0.974	2.166	1.043
Step 2	0.97	8.10	27.83	37.88	25.23	2.217	0.946	2.166	0.998
Step 3	0.99	8.24	28.04	37.82	24.91	2.226	0.947	2.176	0.998
Step 4	1.03	8.43	28.34	37.73	24.47	2.238	0.949	2.190	0.998
Test condition 5, Loss = 5 dB, Noise = 32 dBrnC									
Raw Data	27.14	35.71	31.43	5.71	0.00	3.843	0.889	—	—
Step 1	25.24	40.95	27.16	6.17	0.48	3.843	0.889	3.885	0.922
Step 2	26.89	38.13	26.72	7.42	0.84	3.828	0.938	3.885	0.998
Step 3	27.45	38.19	26.35	7.20	0.80	3.843	0.935	3.902	0.998
Step 4	25.42	37.91	27.70	8.02	0.95	3.788	0.945	3.840	0.998
Test condition 6, Loss = 10 dB, Noise = 32 dBrnC									
Raw Data	18.97	39.66	32.76	6.90	1.72	3.672	0.917	—	—
Step 1	19.46	39.06	31.65	8.93	0.89	3.672	0.917	3.700	0.929
Step 2	21.14	36.80	30.60	10.09	1.37	3.662	0.965	3.700	0.998
Step 3	21.46	36.91	30.38	9.91	1.34	3.672	0.963	3.711	0.998
Step 4	27.72	38.22	26.17	7.10	0.79	3.850	0.933	3.910	0.998
Test condition 7, Loss = 20 dB, Noise = 32 dBrnC									
Raw Data	4.69	6.25	48.44	34.38	6.25	2.688	0.864	—	—
Step 1	1.40	14.76	42.65	33.60	7.59	2.688	0.864	2.684	0.826
Step 2	3.44	17.24	36.64	30.91	11.77	2.697	0.999	2.684	0.998
Step 3	3.37	17.03	36.53	31.11	11.97	2.687	0.999	2.674	0.998
Step 4	5.44	21.94	38.19	26.39	8.03	2.904	1.007	2.900	0.998

Table XV—Continued

Step in the Fit Process	Percentage of Votes					Mean Opinion Score*	Standard Deviation*	Fit Mean	Fit Sigma
	Excel.	Good	Fair	Poor	Unsat.				
Test condition 8, Loss = 30 dB, Noise = 32 dBrnC									
Raw Data	1.92	7.69	15.38	51.92	23.08	2.135	0.920	—	—
Step 1	0.67	6.66	25.95	38.87	27.85	2.135	0.920	2.076	0.981
Step 2	0.76	6.92	25.87	38.26	28.19	2.138	0.931	2.076	0.998
Step 3	0.75	6.87	25.78	38.27	28.33	2.135	0.930	2.072	0.998
Step 4	0.45	4.89	21.72	38.15	34.80	1.980	0.895	1.890	0.998
Test condition 9, Loss = 5 dB, Noise = 42 dBrnC									
Raw Data	11.32	11.32	35.85	33.96	7.55	2.849	1.088	—	—
Step 1	6.83	20.85	34.25	26.55	11.51	2.849	1.088	2.839	1.116
Step 2	4.80	20.59	37.91	27.72	8.98	2.845	1.005	2.839	0.998
Step 3	4.84	20.67	37.93	27.63	8.92	2.849	1.005	2.843	0.998
Step 4	5.23	21.50	38.11	26.83	8.34	2.884	1.006	2.880	0.998
Test condition 10, Loss = 10 dB, Noise = 42 dBrnC									
Raw Data	17.53	17.53	32.99	24.74	7.22	3.134	1.181	—	—
Step 1	14.29	24.81	30.53	20.76	9.60	3.134	1.181	3.150	1.265
Step 2	8.81	27.48	37.97	20.83	4.91	3.144	1.005	3.150	0.998
Step 3	8.63	27.24	38.02	21.07	5.03	3.134	1.006	3.139	0.998
Step 4	7.45	25.51	38.30	22.83	5.90	3.058	1.007	3.060	0.998
Test condition 11, Loss = 20 dB, Noise = 42 dBrnC									
Raw Data	4.35	10.87	19.57	41.30	23.91	2.304	1.081	—	—
Step 1	2.92	11.45	26.35	31.67	27.61	2.304	1.081	2.217	1.206
Step 2	1.11	8.82	28.91	37.54	23.62	2.262	0.953	2.217	0.998
Step 3	1.25	9.51	29.86	37.16	22.23	2.304	0.960	2.263	0.998
Step 4	1.24	9.46	29.79	37.18	22.32	2.301	0.959	2.260	0.998
Test condition 12, Loss = 30 dB, Noise = 42 dBrnC									
Raw Data	0.00	0.00	25.00	0.00	75.00	1.500	0.866	—	—
Step 1	0.98	3.30	9.20	17.76	68.75	1.500	0.866	0.705	1.626
Step 2	0.01	0.25	3.35	17.68	78.71	1.252	0.521	0.705	0.998
Step 3	0.05	1.06	8.84	28.93	61.12	1.500	0.705	1.218	0.998
Step 4	0.12	1.98	13.00	33.70	51.20	1.661	0.786	1.470	0.998

* Mean opinion score and standard deviation are calculated from the percentage of votes given in each table for the corresponding step in the fit process.

respectively, for the MH, HO1, and HO2 tests.

$$\mu_{\text{MH}} = 11.54 - 0.1099|L_e - 11.7| - 0.168N_1 - 0.001059L_eN_1$$
$$\sigma = 1.44, \tag{9}$$

$$\mu_{\text{HO1}} = 7 - 0.1365|L_e - 10.31| - 0.1219N_2 + 0.001577L_eN_2$$
$$\sigma = 1.02, \tag{10}$$

$$\mu_{\text{HO2}} = 7.17 - 0.1681|L_e - 6.7| - 0.1058N_3 + 0.002106L_eN_3$$
$$\sigma = 0.998, \tag{11}$$

where

L_e = Acoustic-to-acoustic loudness loss (in dB) of an overall telephone connection, determined using the Electro-Acoustic Rating System (EARS) method.

N = Circuit noise (in dBrnC) at the input to a set with a receive-loudness rating of 26 dB, determined using the EARS method.

N_1 = Total noise in dBrnC resulting from power addition of the circuit noise, N, from the MH tests with 34.03 dBrnC.

N_2 = Total noise in dBrnC resulting from power addition of the circuit noise, N, from the HO1 tests with 23.76 dBrnC.

N_3 = Noise, N from the HO2 tests.

The values 34.03 and 23.76 were determined as fit parameters. The particular functional form was selected to provide as simple a model as possible of the systematic effects observed in the data.

The results represented by eqs. (9), (10), and (11) revealed two important differences between the MH and HO tests. First, the standard deviation, σ, was considerably larger for the MH tests than for either HO1 or HO2. Second, the subjective opinions, as represented by the means, μ, calculated from eqs. (9), (10), and (11) were considerably higher in the MH test compared with the HO tests. These differences occurred despite the similarities of the tests. A careful examination of either the raw data or the smooth results clearly shows that the subjects' ratings tended to be more critical in the two HO tests compared with the subjects in the MH test.

A clearer picture of the differences is obtained by selecting a set of loss (L_e) and circuit noise (N) values over a common range of the tests for MH, HO1, and HO2 and computing the corresponding values of the means (μ) for the three tests from eqs. (9), (10), and (11). If plots are made of the MH means versus both the HO1 and HO2 means and the appropriate linear regression made for both plots, then eqs. (12) and (13), respectively, represent the regression line between the MH and HO1 means and the MH and HO2 means. Such a plot is shown in Fig. 20 for MH and HO1.

$$\mu_{\mathrm{MH}} = 1.372\mu_{\mathrm{HO1}} - 0.206. \tag{12}$$

$$\mu_{\mathrm{MH}} = 1.288\mu_{\mathrm{HO2}} - 0.215. \tag{13}$$

The Pearson product moment coefficient of correlation was found to be 0.9586 and 0.9693 for MH with HO1 and HO2, respectively. Eqs. (12) and (13) clearly show the difference between the means for the MH and the two HO tests. These equations also show the close agreement between the two Holmdel tests.

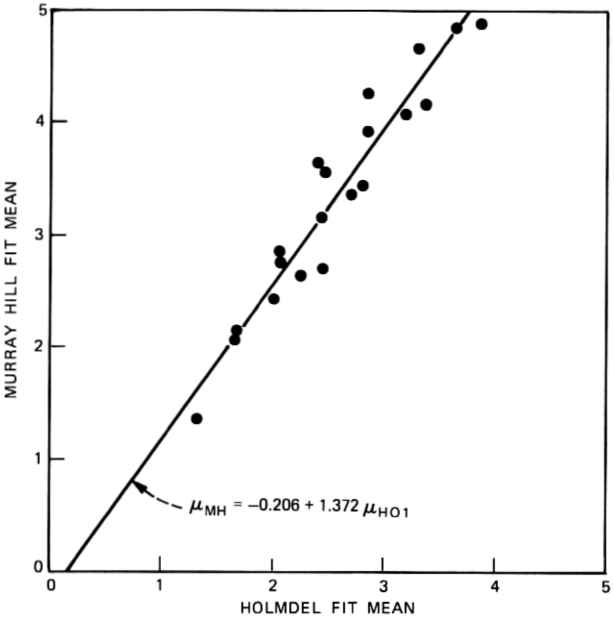

Fig. 20—Comparison of 1965 Murray Hill and 1969 Holmdel 1 loss-noise results.

Equations (12) and (13) are used to adjust for the difference between the HO results and the MH results in Section VI of this paper.

5.3 Echo analysis

The four echo subjective tests yielded data on the subjective effects of echo-path loudness loss, E, in dB and echo-path delay, D, in ms. Circuit noise, N, was not the same for all four tests, and needed to be considered as a test variable. Loudness loss, L_e, was a factor in only two of these tests (the other two were listening only tests) and, as a first approximation, it was decided to ignore L_e and concentrate on E, D, and N in the analysis.

Preliminary analyses of the individual test data indicated that there were only relatively small differences in the absolute ratings among the 1968–1970 tests. Thus, it was feasible to combine these data and use the analysis method described previously. The resulting equation relating the normal density means to the test variables, as realized at Step 4 in the analysis, was a function of E, D, and N, where N for this preliminary analysis was the actual noise at the telephone set terminals. The effect of noise was asymptotic. That is, the fit mean, μ, was determined largely by the values of E and D when echo was the predominant impairment, while, for any value of D, increasing the value of E gradually led to the mean being solely determined by the value of N. This asymptotic effect of the noise made it relatively simple to separate the effects of noise and echo in the function depicting the mean as a function of the test variables. The resulting functions are given in eq. (14).

$$\begin{aligned}
\mu_{E1} &= 4.74 - 3.32 \log_{10} [(1 + D)/\sqrt{1 + (D/480)^2}] + 0.1414E \\
\mu_N &= 6.38 - 0.094N \\
\mu_{NE} &= \frac{\mu_E + \mu_N}{2} - \sqrt{\left(\frac{\mu_E - \mu_N}{2}\right)^2 + (0.627)^2} \qquad (14) \\
\sigma &= 0.888.
\end{aligned}$$

Although these functions were a useful interim result, subsequent analysis described in the later sections of this paper indicated that when the effects of both loss and noise were included and the noise was referred to the input of a reference set, further modifications were necessary. These functions are included here because they provided a basis for the subsequent modifications.

VI. COMBINATION OF MODELS

The results of the three loss-noise tests as given by eqs. (9), (10), and (11) showed fundamental differences among the tests. Despite the similarity of the tests and the general character of the results, both the raw data and the smoothed results showed that the subjects' ratings in the two HO tests tended to be more critical in their evaluations than they were in the MH tests. This could have occurred because of one or more differences in the tests, such as room noise, sidetone path loss, year of test, or some fundamental difference in the attitude of the subject groups. The frequent repetition of high-quality conditions in the Holmdel tests may also have been a factor.

The exact reasons for the differences in the test results could not be determined. Because of these differences, direct pooling of results from the three tests did not appear to be justifiable. However, the test results were combined by adjusting the HO1 and HO2 results to a MH base using the linear transformation obtained from the linear regressions introduced previously in eqs. (12) and (13). In this way, the systematic differences among the test results were preserved, while achieving the advantages of a larger data base. The transformed means were then included with the MH means, and a new equation was obtained by applying Step 4 to this combined set of means. Table XVI shows the fit means at Step 3 for the MH, HO1, and HO2 test results and the adjusted fit means for the HO1 and the HO2 test results. Thus, the final fit was based on a total of 51 conditions. The final result

Table XVI — Data for combined fit—Step 3

Murray Hill			Holmdel 1				Holmdel 2			
Connection Loudness Loss (dB)	Circuit Noise (dBrnC)	Fit Mean (μ_a)	Connection Loudness Loss (dB)	Circuit Noise (dBrnC)	Fit Mean (μ_a)	Adjusted Fit Mean (μ_a')	Connection Loudness Loss (dB)	Circuit Noise (dBrnC)	Fit Mean (μ_a)	Adjusted Fit Mean (μ_a')
5	21	4.53	—	—	—	—	5	25	4.48	5.56
10	21	5.08	10	22	4.18	5.53	10	25	4.51	5.59
15	21	5.23	15	22	3.94	5.20	—	—	—	—
20	21	4.27	20	22	3.35	4.39	20	25	3.33	4.07
25	21	3.62	25	22	2.36	3.03	—	—	—	—
30	21	2.90	30	22	2.46	3.17	30	25	2.18	2.59
5	28	4.80					5	32	3.90	4.81
10	28	5.39					10	32	3.71	4.56
15	28	4.80					—	—	—	—
20	28	3.56					20	32	2.67	3.22
25	28	3.41					—	—	—	—
30	28	2.29					30	32	2.07	2.45
5	36	4.59	—	—	—	—				
10	36	5.01	10	35	3.17	4.14				
15	36	4.01	—	—	—	—				
20	36	3.05	20	35	2.54	3.28				
25	36	2.20	25	35	1.94	2.46				
30	36	2.09	30	35	1.86	2.35				
			10	40	2.97	3.87				
			20	40	1.73	2.17				
			25	40	1.72	2.15				
5	44	3.24	—	—	—	—	5	42	2.84	3.44
10	44	2.93	10	45	2.06	2.62	10	42	3.14	3.83
15	44	3.21	—	—	—	—	—	—	—	—
20	44	2.16	20	45	1.76	2.21	20	42	2.26	2.70
25	44	1.91	—	—	—	—	—	—	—	—
30	44	0.54	30	45	0.93	1.07	30	42	1.22	1.36

is eq. (15).

$$\mu_{\mathrm{MH}} = 10.36 - 0.185\sqrt{(L_e - 7.2)^2 + 1} - 0.1647N_F + 0.00167L_E N_F, \quad (15)$$

where

N_F = Power addition of noise with 27.37 dBrnC.

6.1 General rating scale

Equation (15) above is calculated in terms of the MH base. This equation together with eqs. (12) and (13) can be used to express the representation of the subjective ratings in terms of any one of the other test bases, HO1 or HO2. Each test base has a standard deviation associated with it that can then be used in conjunction with the computed means to calculate predicted vote histograms from the normal density curves. However, if this is done, the fit means and vote histograms will be different for each test in accordance with the difference in absolute ratings obtained for each test. To eliminate the need for three separate equations, one for each test, a general transmission-rating scale was established.

This transmission rating scale, referred to as the R-scale, is simply a linear transformation of the normal density means, defined by eq. (15), with the constraints that two preselected transmission conditions are to be the anchor conditions for the transformation. R-scale values of 80 and 40, respectively, were selected for the transmission conditions $L_e = 15$, $N = 25$, and $L_e = 30$, $N = 40$. These two transmission conditions were selected to be well separated in quality. The first pair is typical of a short intertoll connection and the latter represents an extreme condition of loss and noise that should rarely occur even on long intertoll connections between long loops.

Using the above transmission conditions as anchors, R-scale values can be specified in terms of μ for each test through the linear transformation $R = a + b\mu$, with a and b determined from the anchor constraints.

6.2 Loss-noise model

From eq. (15), the transmission condition $L_E = 15$, $N = 25$ yields $\mu_{\mathrm{MH}} = 4.806$ and $L_e = 30$, $N = 40$ yields $\mu_{\mathrm{MH}} = 1.528$. Using these values of μ, respectively, with R-scale values of 80 and 40 determines the transformation to the R-scale from the μ scale as given in eq. (16).

$$R = 21.37 + 12.20\mu_{\mathrm{MH}}. \quad (16)$$

Substituting eq. (15) for μ_{MH} into eq. (16) gives eq. (17), which is

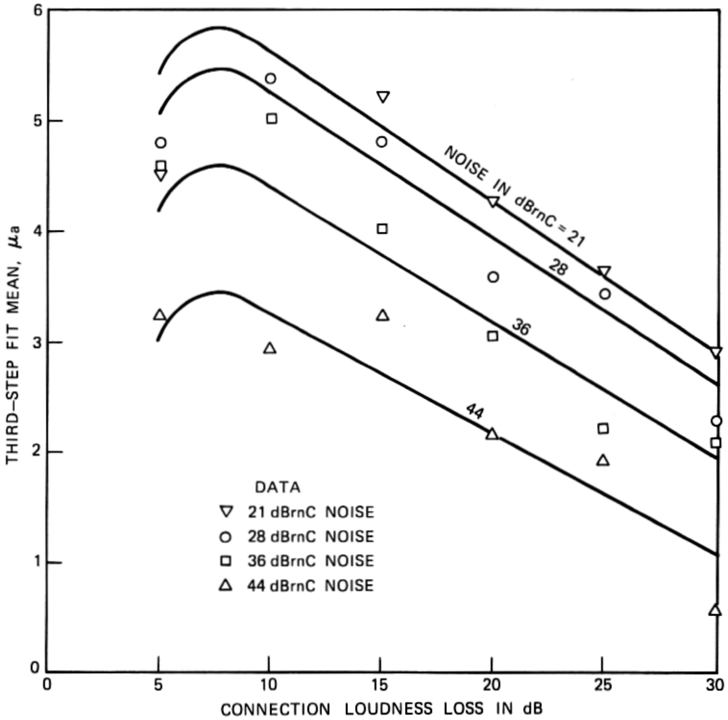

Fig. 21—MH loss-noise test, Step 3 means from data as a function of loudness loss and noise compared with means predicted from the loss-noise model at the MH base.

the R-scale representation of the subjective opinion for loss and noise.

$$R_{LN} = 147.76 - 2.257\sqrt{(L_e - 7.2)^2 + 1} - 2.009N_F + 0.02037L_eN_F. \tag{17}$$

Equation (17) is plotted in Fig. 1 as transmission rating versus L_e for a selected set of values of N. These curves represent the predicted transmission rating, in terms of the R-scale, for selected values of L_e and N.

The R-scale result of eq. (17) can also be used in conjunction with the appropriate standard deviations associated with eqs. (9), (10), and (11) and the appropriate inverse linear regression lines of eqs. (12) and (13) to calculate percent good or better, poor or worse, or other characteristics at the chosen test base.

For the μ-scale, the proportion of ratings good or better is the integral of the standard normal density curve from $(3.5 - \mu)/\sigma$ to infinity. In the R-scale, this corresponds to the integral of the standard normal density curve from minus infinity to $[R - (a + 3.5b)]/\sigma b$.

Similar computations can be made for proportion poor or worse, or for proportions of ratings in any of the five categories. The appropriate limits of integration to compute the proportion of ratings good or better and poor or worse are given in Table II for the three loss-noise test bases (MH, HO1, HO2).

The above discussion concerning the relationships between the proportions good or better and poor or worse and the R-scale lead to the plots of Fig. 4 which show these relationships for the three test bases.

Finally, the results summarized in Table II were also used to generate curves showing the tradeoff between L_e and N for selected values of percent good or better and percent poor or worse as shown in Figs. 6 and 7, respectively. As noted on the figures, these results correspond to the MH data base.

In Figs. 21 to 23, the third-step fit means for the individual tests are plotted as a function of loudness loss with circuit noise as a parameter. The solid lines in the figure correspond to the values of transmission rating in the final model transformed by the appropriate relation between μ and R for each test. These figures show a generally good fit to the individual test results for each of the tests.

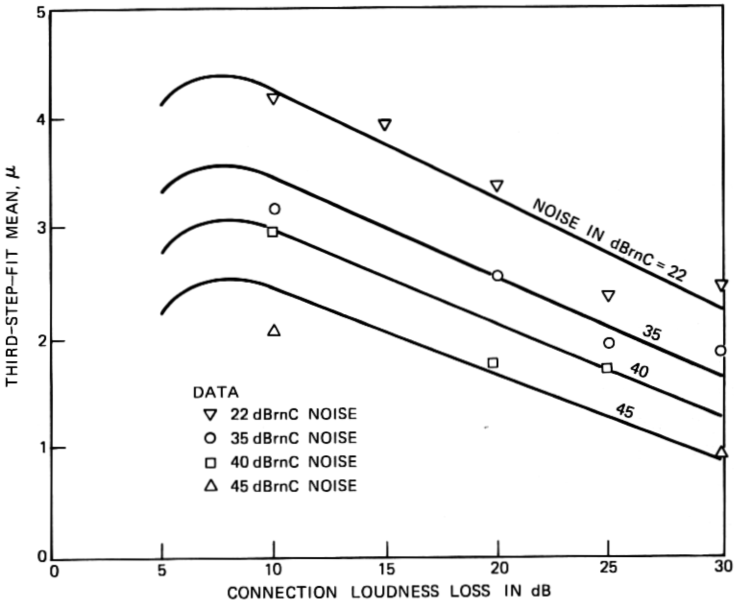

Fig. 22—HO1 loss-noise test, Step 3 means from data as a function of loudness loss and noise compared with means predicted from the loss-noise model at the HO1 base.

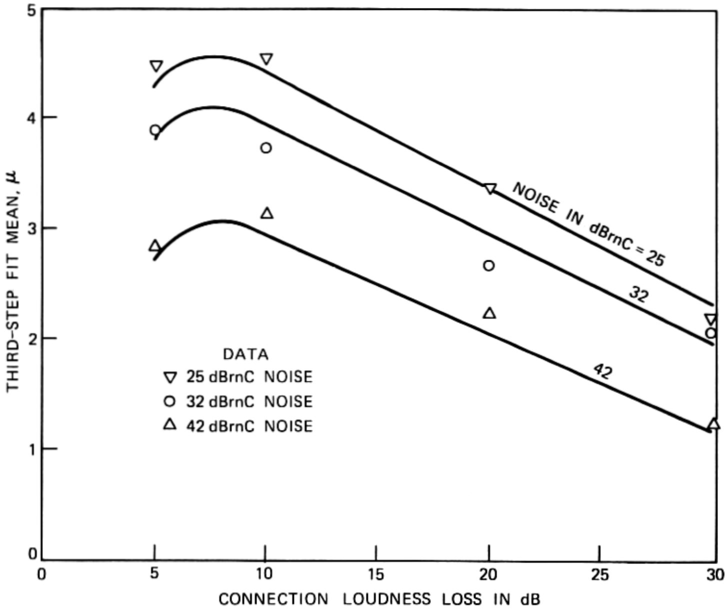

Fig. 23—HO2 loss-noise test, Step 3 means from data as a function of loudness loss and noise compared with means predicted from the loss-noise model at the HO2 base.

6.3 Talker-echo model

The R-scale was introduced previously in terms of L_e and N. Thus, it was necessary to use the conversational echo test results which included both loss and noise as parameters to establish an expression for the echo results on the R-scale. The 1970 SIBYL test was used as a basis for this conversion. As indicated in Table XIV, this test included a base condition with $L_e = 10$ dB and $N = 30$ dBrnC for which the fit mean, μ, was 4.01. The R-scale value corresponding to this combination of loss and noise is 83.47. Similarly, the results for the noise condition in the 1968 lab tests were taken into account but given less weight because they were obtained in a less realistic test environment. A further aid in deriving this conversion was the HO1 SIBYL test for loss and noise which preceded the 1970 SIBYL echo test and was conducted in the same manner with the same subjects. The final conversion is given in eq. (18) below.

$$R = 18.7 + 16.1\mu_{E1}. \tag{18}$$

This relationship is almost identical to that of the HO1 SIBYL tests at low values of μ. However, a slight correction was included for higher values of μ to take account of the base condition in the 1970 SIBYL echo test. For the HO1 test the value of μ corresponding to the base condition ($R = 83.47$) was 3.86. The relationship in eq. (18) gives a value of 4.02 and provides good agreement with the actual value of μ which was 4.01 in the 1970 SIBYL echo tests.

Substituting eq. (14) for μ_{E1} into eq. (18) yields eq. (19) which is the transmission-rating model for echo.

$$R_E = 95.01 - 53.45 \log_{10} [(1 + D)/\sqrt{1 + (D/480)^2}] + 2.277E. \quad (19)$$

Equation (19) is plotted in Fig. 2.

In the analysis above, eq. (19) was derived from eqs. (14) and (18) to provide excellent agreement for the 1970 SIBYL echo tests. However, these same relationships did not provide good agreement with the 1970 lab tests which had a higher loss included in all conditions. The base condition for this test as given in Table XII, with $L_e = 18$ dB and $N = 33$ dBrnC ($R = 67.36$), had a fit mean, μ, of 3.59.

The equation,

$$R = 20 + 13.33\mu_{E2}, \quad (20)$$

provided a good match at this point and retained the relationship between μ and R at low values of μ obtained previously for the SIBYL echo tests. With the relationship defined by eq. (20), the transmission-rating model for echo given in eq. (19) provided an excellent representation of the results from the 1970 lab test for echo. The extent of the agreement is illustrated in Section 6.4 where the combined loss-noise-echo model is discussed.

6.4 Loss-noise-echo model

In the development of the echo results, it was noted that the degradations considered were echo-path-loudness loss (E), echo-path delay (D), loudness loss (L_e), and noise (N). Loss and noise were eliminated in the final echo result because it was felt that for any combined result, the loss and noise influence should be based on the larger data base available from the SIBYL tests.

The original analysis of the echo-loss-noise data showed that the loss and noise really only affected the circuit performance as an asymptote. That is, for very large E, the ratings were determined by L_e and N. Use was made of this fact when the echo result of eq. (14) was developed. The combination of R_{LN} and R_E was made such that this asymptotic behavior was retained in the final model.

The final result for loss, noise, and echo is presented in terms of the R-scale as R_{LNE}. This final result is shown in eq. (21).

$$R_{LNE} = \frac{R_{LN} + R_E}{2} - \sqrt{\left(\frac{R_{LN} - R_E}{2}\right)^2 + C^2}. \quad (21)$$

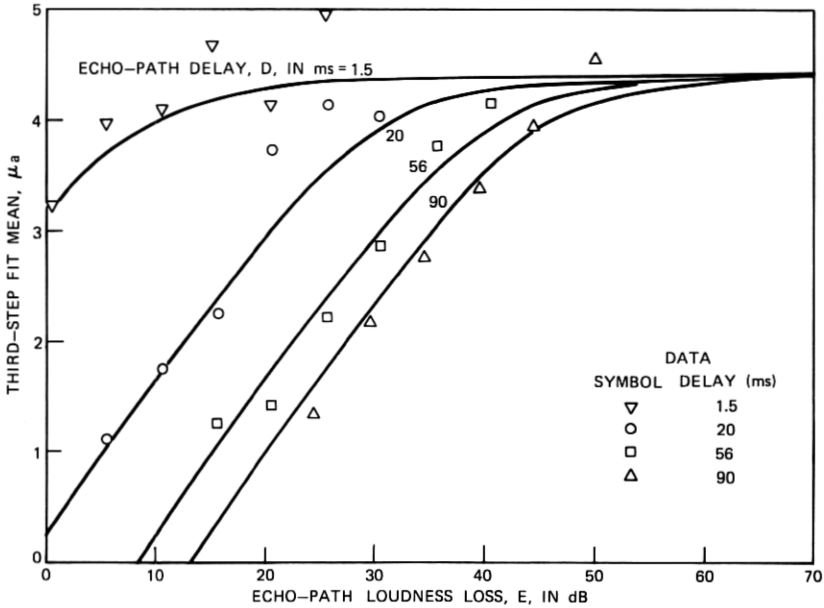

Fig. 24—1968 laboratory echo test (noise=18 dBrnC, loss=10 dB) Step 3 means from data as a function of echo-path loss and delay compared with means predicted from the loss-noise-echo model at the echo 1 base.

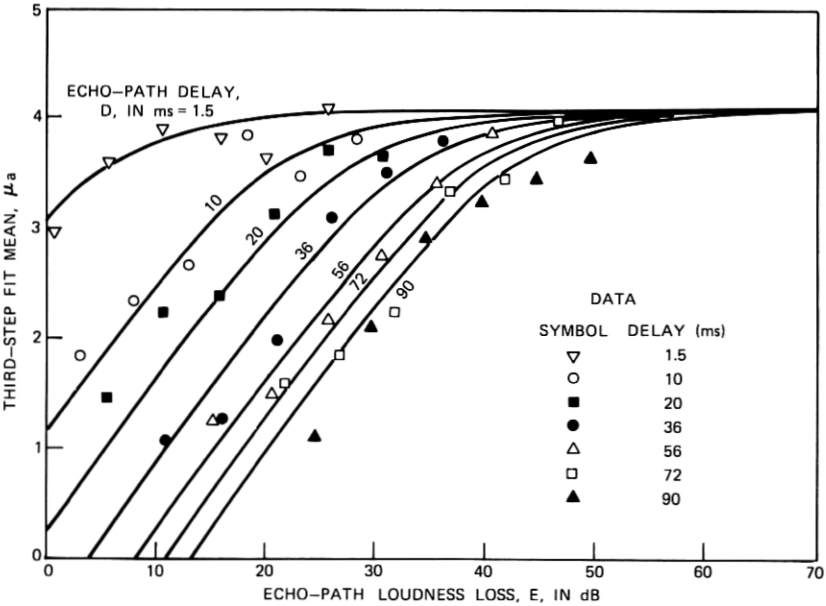

Fig. 25—1968 laboratory echo test (noise=28 dBrnC, loss=10 dB) Step 3 means from data as a function of echo-path loss and delay compared with means predicted from the loss-noise-echo model at the echo 1 base.

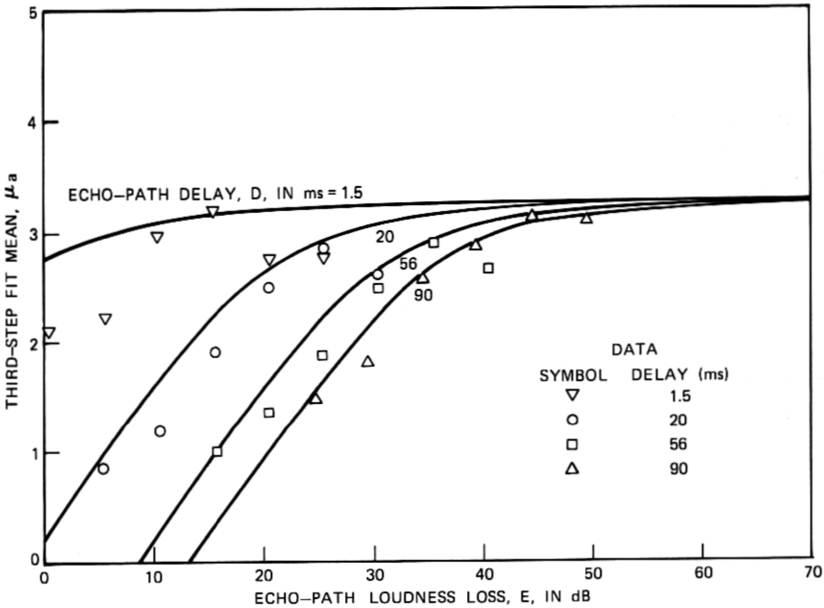

Fig. 26—1968 laboratory echo test (noise=38 dBrnC, loss=10 dB) Step 3 means from data as a function of echo-path loss and delay compared with means predicted from the loss-noise-echo model at the echo 1 base.

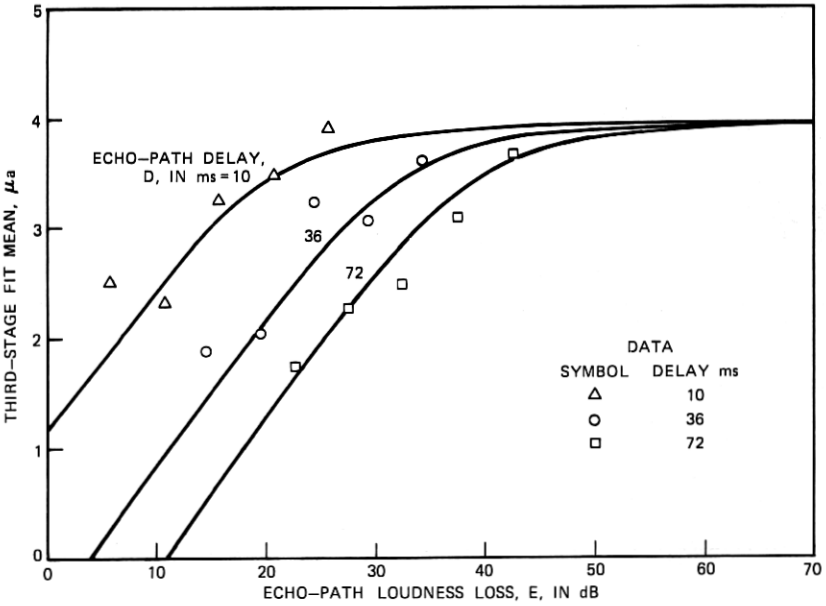

Fig. 27—SIBYL echo test, Step 3 means from data as a function of echo-path loss and delay compared with means predicted from the loss-noise-echo model at the echo 1 base.

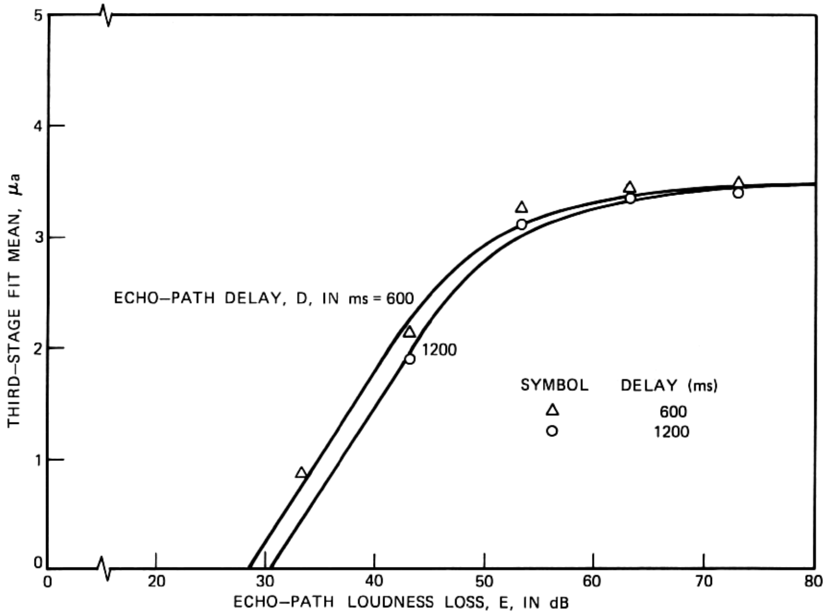

Fig. 28—1970 laboratory echo test, Step 3 means from data as a function of echo path loss and delay compared with means predicted from the loss-noise-echo model at the echo 2 base.

With $C = 0$, R_{LNE} is simply the lesser of R_{LN} and R_E. The factor C is selected to represent the additional degradation when R_{LN} and R_E are nearly equal. The value of $C = 10$ was based on echo tests that included echo, loss, and noise and was obtained as the product of the constant, 0.627, in eq. (14) and the slope of the line relating R and μ given in eq. (18).

The final result in terms of the R-scale is:

$$R_{LNE} = \frac{R_{LN} + R_E}{2} - \sqrt{\left(\frac{R_{LN} - R_E}{2}\right)^2 + (10)^2}. \tag{22}$$

For high echo-path-loudness loss, the ratio is determined mainly by connection loudness loss and circuit noise and the result reduces to the R_{LN} result. Similarly, for connection loudness loss near optimum and low circuit noise, the rating is determined mainly by the echo, and the result effectively reduces to the R_E result.

Comparison of the final model for loss, noise, and echo with the third-step-fit means for the individual tests are shown in Figs. 24 to 28. As in the case of the loss-noise model, the final loss-noise-echo model provides good agreement with the test results from each of the individual tests.

VII. ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of a large number of colleagues who contributed to the studies described in this paper. Particular note is due Mr. T. K. Sen who was principally responsible for the loss-noise subjective tests and to S. H. Franz, J. T. Powers, D. J. Rhoads, and E. J. Thomas who conducted the talker-echo tests.

APPENDIX A

Examples Demonstrating Use of the Transmission-Rating Models

A.1 Example 1—local connection

500-type telephone sets on a calling and a called customer loop, each of which consists of 1 kilofoot of 26-gauge nonloaded cable:

Loss (L_e): $\text{TLR} = -23.6$ dB
$\text{RLR} = 26.3$ dB
$L_e = (\text{TLR} + \text{RLR})$ dB
$= 2.7$ dB

Noise (N): N_T = Total noise at each telephone set from two loops, each of which meet the 20 dBrnC loop-noise objective.[5]
= 23 dBrnC
N = Total noise referred to a telephone set with RLR of 26 dB
= 27 dBrnC.

Talker Echo (assumed negligible):

Using Table II with $L_e = 2.7$ dB and $N = 27$ dBrnC,

$$R_{LN} = 78.3.$$

Using Table II for the MH data base,

$$GoB = 79.2\%, \qquad PoW = 6.6\%.$$

A.2 Example 2—local connection

500-type telephone sets and a calling and a called customer loop, each of which consists of 8 kilofeet of 26-gauge nonloaded cable:

Loss (L_e): $\text{TLR} = -18.9$ dB
$\text{RLR} = 27.3$ dB
$L_e = 8.4$ dB

Noise (N): Assumed to be the same as for Example 1:
$N = 27$ dBrnC.

Talker echo (assumed negligible):

Using Table II with $L_e = 8.4$ dB and $N = 27$ dBrnC,

$$R_{LN} = 88.7.$$

Using Table II for the MH data base,

$$GoB = 92\%, \qquad PoW = 1.8\%.$$

A.3 Example 3—local connections

500-type telephone sets on a calling and a called customer loop, each of which consists of 15 kilofeet of 26-gauge nonloaded cable:

Loss (L_e): TLR = −13.2 dB
RLR = 30.1 dB
L_e = 16.9 dB

Noise (N): Assumed to be the same as for Example 1.

$$N = 27 \text{ dBrnC}.$$

Talker echo (assumed negligible):

Using Table II with $L_e = 16.9$ dB and $N = 27$ dBrnC,

$$R_{LN} = 75.5.$$

Using Table II for the MH data base,

$$GoB = 74.2\%, \qquad PoW = 8.9\%.$$

A.4 Example 4—toll connection

500-type telephone sets on a calling and a called customer loop, each of which consists of 8 kilofeet of 26-gauge nonloaded cable:

Loss (L_e): TLR = −18.9 dB
RLR = 27.3 dB
L = Class 5 office-to-Class 5 office loss
= 7.7 dB (mean for the connection-length category, 775 to 2900 miles, Table VI of Ref. 12)
L_e = TLR + RLR + 7.7
= 16.1 dB.

Noise (N): Assume no noise from the loops:

N_T = Total noise from the Class 5 office-to-Class 5 office connection (33.8 dBrnC for the connection length category 775 to 2900 miles, Table III of Ref. 12 referred to the telephone set)
= 28.5 dBrnC
N = 32.5 dBrnC.

Talker echo (assumed negligible):

Using Table II with $L_e = 16.1$ dB and $N = 32.5$ dBrnC,

$$R_{LN} = 71.$$

Using Table II for the MH data base,

$$GoB = 65.2\%, \qquad PoW = 13.9\%.$$

A.5 Example 5—toll connection

Same as Example 4, Section A.5, except that it takes into account talker echo:

Echo-path delay (D)
D = Far-end echo-path loss
= 37.3 ms (mean for connection length 1450 to 2900 miles, Table III of Ref. 11).

Echo-path loudness loss (E)
E' = Loss of echo path from near-end Class 5 office to distant end and return,
= 23.3 dB (mean for connection length 1450 to 2900 miles, Table II of Ref. 11).
E = TLR + RLR + E'
= 31.7 dB.

Using Table II with $L_e = 16.1$ dB, $N = 32.5$ dBrnC, $D = 37.3$ ms, and $E = 31.7$ dB.

$$R_{LN} = 71$$
$$R_E = 82.6$$
$$R_{LNE} = 65.2.$$

Using Table II for the MH data base,

$$GoB = 52.6\%, \qquad PoW = 22.4\%.$$

REFERENCES

1. T. C. Spang, "Loss-Noise Echo Study of the Direct Distance Dialing Network," B.S.T.J., *55* (January 1976), pp. 1–36.
2. H. R. Huntley, "Transmission Design of the Intertoll Telephone Trunks," B.S.T.J., *32* (September 1953), pp. 1019–1036.
3. G. M. Phillips, "Echo and Its Effects on the Telephone User," Bell Laboratories Record (August 1954), pp. 281–284.
4. O. H. Coolidge and G. C. Reier, "An Appraisal of Received Telephone Speech Volume," B.S.T.J., *38* (May 1959), pp. 877–897.
5. D. A. Lewinski, "A New Objective for Message Circuit Noise," B.S.T.J., *43* (March 1964), pp. 719–740.
6. T. K. Sen, "Subjective Effects of Noise and Loss in Telephone Transmission," IEEE Trans. Commun. Technol., *COM-19* (December 1971), pp. 1229–1233.
7. A. H. Inglis and W. L. Tuffnell, "An Improved Telephone Set," B.S.T.J., *30* (April 1951), pp. 239–270.

8. J. L. Sullivan, "A Laboratory System for Measuring Loudness Loss of Telephone Connections," B.S.T.J., *50* (October 1971), pp. 2663–2739.
9. A. J. Aikens and D. A. Lewinski, "Evaluation of Message Circuit Noise," B.S.T.J., *39* (July 1960), pp. 879–909.
10. W. T. Cochrane and D. A. Lewinski, "A New Measuring Set for Message Circuit Noise," B.S.T.J., *39* (July 1960), pp. 911–931.
11. F. P. Duffy et al., "Echo Performance of Toll Telephone Connections in the United States," B.S.T.J., *54* (February 1975), pp. 209–243.
12. F. P. Duffy and T. W. Thatcher, Jr., "Analog Transmission Performance on the Switched Telecommunications Network," B.S.T.J., *50* (April 1971), pp. 1311–1347.
13. H. D. Irvin, "Studying Tomorrow's Communications ··· Today," Bell Laboratories Record (November 1958), pp. 398–402.
14. H. D. Irvin, "SIBYL: A Laboratory for Simulation Studies of Man-Machine Systems," 1958 IRE Wescon Conv. Rec., Part 4, pp. 277–285.
15. J. L. Sullivan, "Is Transmission Satisfactory? Telephone Customers Help Us Decide," Bell Laboratories Record (March 1974), pp. 90–98.
16. A. M. Noll, "Subjective Effects of Sidetone During Telephone Conversation," Commun. Elec., *83* (May 1964), pp. 228–231.
17. A. Michael Noll, "Effects of Head and Air-Leakage Sidetone During Monaural-Telephone Speaking," J. Acoust. Soc. Amer., *36* (March 1964), pp. 598–599.
18. American National Standards Institute, *Specification for Sound Level Meters, S1.4-1971*.
19. International Telecommunication Union, *Telephone Transmission Quality, Local Networks and Telephone Sets*, CCITT Green Book, *5*, 1973.
20. R. D. Prosser, J. A. Allnatt, and N. W. Lewis, "Quality Grading of Impaired Television Pictures," Proc. IEE., *111*, No. 3 (March 1969), pp. 491–502.