5i’

PART II

SUPPLEMENTS TO SERIES P RECOMMENDATIONS

MONTAGE: PAGE PAIRE = PAGE BLANCHE

Supplement No. 1

PRECAUTIONS TO BE TAKEN FOR CORRECT INSTALLATION AND MAINTENANCE OF AN IRS

(For this Supplement see Volume V of the *Orange Book*)

Supplement No. 2

METHODS USED FOR ASSESSING TELEPHONY TRANSMISSION PERFORMANCE

(Geneva, 1980; modified at Malaga-Torremolinos, 1984; Melbourne, 1988)

(Quoted in Recommendation P.80)

(Contribution from British Telecom)

1 Introduction

This Supplement gives brief descriptions of the methods for assessing telephony transmission performance that are recommended by the CCITT or have been employed over Study Periods 1968 to 1980 in studying Questions assigned to Study Group XII. Some of the methods are already fully described in Recommendations and these will merely be listed here with reference to the appropriate Recommendation. Other methods are also described in detail elsewhere; the essential features of these are given here with a brief description of how they are conducted, with reference to descriptions published elsewhere.

2 List of methods

a) loudness comparison for speech (reference equivalents and loudness ratings);

b) articulation (AEN) ratings;

c) listening opinion tests;

d) conversation opinion tests;

e) quantal-response detectability tests.

3 Brief descriptions and references to more complete descriptions

3.1 Loudness comparisons for speech are intended to quantify the relative level at which speech, transmitted over a given telephone connexion, reaches the ears of customers while they are listening to a person talking at the other end. In order to standardize the measuring procedure, the talking and listening conditions are each controlled in a specified manner. Circuit noise and room noise are excluded from the determination and so the results are governed by the overall mouth-to-ear transmission loss of the speech path being considered. The present recommended method is given in Recommendation P.72 (*Red Book*) and proposals for new methods are to be found in Question 15/XII [1]. More general information can be found in Reference [2].

3.2 Articulation measurements are based on measurement of the fraction of speech sounds recognized correctly when transmitted and reproduced over the speech path in question. Circuit noise and room noise at specified levels should be present and the result is affected by their levels. Just as for § 3.1 above, talking and listening conditions are controlled. The method recommended by the CCITT is described in former Recommendation P.45 (*Orange Book*). Other information will be found in Reference [2].

3.3 Listening opinion tests are conducted using speech material in the form of sentences and the listeners judge the speech received over the path according to a given criterion. The method has been widely used, and further details can be found in Reference [2].

3.3.1 *Method of conducting listening opinion tests*

The speech is usually recorded so that it can be reproduced at a given level. The recordings for this purpose must be carefully made and copied so that uncontrolled degradations do not appear. Circuit noise and room noise may be present, and their effects are taken into account.

Two subjective criteria commonly used are loudness preference and listening effort, for which the following scales are used.

— Loudness preference scale :

Opinion scale No. 4A

A Much louder than preferred.

B Louder than preferred.

C Preferred.

D Quieter than preferred.

E Much quieter than preferred.

— Listening effort scale :

Opinion scale No. 7: Opinions based on the effort required to understand the meanings of sentences

A Complete relaxation possible; no effort required.

B Attention necessary; no appreciable effort required.

C Moderate effort required.

D Considerable effort required.

E No meaning understood with any feasible effort.

The heading ‘‘. | | *Effort required to understand the meanings of sentences* ’’ is particularly important. Without it, the other descriptions are liable to be seriously misunderstood.

Experimental design is usually based on a graeco-latin or hyper-graeco-latin square, in which rows represent listeners, columns represent the order in which conditions are administered, symbols of the first alphabet represent circuit conditions, and symbols of other alphabets represent talkers and lists of sentences. Each cell of the design thus represents a ‘‘run’’, in which a particular list of sentences, recorded by a particular talker, is replayed via a particular circuit condition to a particular listener at a particular position in the sequence of conditions presented to that listener. Within each run the listening level is varied over a number of predetermined values in random order, one value per group of five sentences, and the subject votes on one of the above opinion scales at the end of each group. Rarely some other parameter, such as bandwidth, is varied within each run instead of listening level.

In listening-effort tests, listeners are specially prone to what is known as the ‘‘enhancement’’ effect: that is, their standards of judgement are liable to be strongly influenced by the range of quality and listening level occurring in the same test, and especially within the same run. It is therefore important that the circuit conditions chosen should not include too many bad ones (that is, conditions that will yield a poor listening-effort score even with the best listening levels), that every run should cover a range of listening levels from well above optimum to at least 30 dB below optimum, and that within each run at least one group of sentences should be heard via an ‘‘anchor’’ condition (a good condition with a good listening level). It is also important that groups and lists of sentences should not vary too widely in their intrinsic comprehensibility, and that no subject should hear the same sentence more than once in the same experiment, because the listening effort needed to understand a familiar sentence would obviously be reduced.

The votes using the above scales are scored respectively 4, 3, 2, 1 and 0: the mean of these values for each circuit condition is called ‘‘ mean opinion score to verify that the effects due to circuit condition, listening levels, talkers, listeners and other factors are as expected, to determine their significance, and to evaluate confidence intervals. It is usual to express the relationship between listening level and loudness-preference mean opinion score (scale 4A) by fitting an equation describing a straight line or logistic curve, whereas the relationship between listening level and listening-effort mean opinion score (scale 7) is expressed by a fitted quadratic or more complicated equation; other features of the circuit conditions may also enter as parameters into these equations.

Listening tests using sentence material can also be conducted as pair-comparisons, but these should be undertaken with due consideration to ensure that subjects become suitably adapted to each test condition.

3.4 Conversation tests may be conducted either as interviews after real customers have made actual calls or as laboratory tests. Further information regarding methods recommended by the CCITT for the former is given in Recommendation P.82. Laboratory conversation tests are intended as far as possible to reproduce under laboratory conditions the actual service conditions experienced by telephone customers: to this end it is necessary to choose the circuit conditions and subjects suitably, and to administer the tests in an appropriate manner. A method intermediate between field observations and laboratory tests is that used by the AT&T and called SIBYL (see also Reference [3]). Particulars of the method used by British Telecom are given below.

3.4.1 *Method for conducting conversation tests*

The need for careful and exhaustive preparations cannot be too strongly emphasized. It will be obvious to all that the connexions must be correctly specified and set up, and measured accurately before and after each experiment; that auxiliary facilities such as dialling and ringing must be provided, so that any of the desired connections can be chosen and established quickly and without error; and that faithful records of the output of each test must be kept. But some other equally important considerations are less obvious. The following gives an outline of a system that takes all these matters into account, and has been found satisfactory in British Telecom.

3.4.1.1 *Experimental design*

The most suitable designs are of the $n \times n$ | raeco-latin square type, where each of n pairs of subjects carries out one conversation on each of n circuit conditions. Precision is very low if n is less than 8; at the other extreme it is not practical to expect subjects to attend on more than four occasions, or to carry out more than four conversations per visit. Moreover, the total number of conversations, $n \times n$, increases much more rapidly than n. For this reason, n is normally limited to the range 8 to 15 inclusive: graeco-latin squares (with symbols from two alphabets) exist for all these numbers. In such a design, the convention is that rows denote pairs of subjects; columns denote the order of administering the experiment; symbols of the first alphabet denote circuit conditions (distinguished not only according to properties of the connections by themselves, but also according to room noise levels and any other ‘‘treatment’’ factors); symbols of the second alphabet denote sets of pictures used as the topic of conversation. No further

orthogonal factors can be incorporated at all where $n = 10$, 14 or 15, nor by any simple method when $n = 12$; but where $n = 8$, 9, 11 or 13, it is possible to construct hyper-graeco-latin squares with symbols from (n — 3) additional alphabets, which may be used to govern further orthogonal factors (such as selection of carbon microphones, choice of calling party, or choice of crosstalk recording), for each conversation. When the square is not hyper-graeco-latin these factors must be allocated by some simple balanced rotation scheme, but this may give rise to biases that cannot be eliminated from the results. For this reason the recommended value of n is now 13 rather than 12 as previously.

To the basic square is added an extra column at the beginning, having the same circuit condition and the same picture set for all pairs of subjects. This column represents a preliminary conversation for each pair of subjects, which serves to accustom them to the procedure, and to some extent stabilizes their standards of judgement. Thus each of the n pairs of subjects carries out ($n + 1$) conversations altogether. The results from the preliminary conversations are not included in the main part of the analysis of results, but are analyzed separately. Using the same preliminary circuit condition in different experiments establishes some common ground between experiments, but if precise comparisons between results from different experiments are desired, care must be taken to include replications of several standard circuit conditions in each such experiment.

3.4.1.2 *Choice of circuit conditions*

Circuit conditions between which particularly precise comparisons are desired must be included within the same experiment.

Besides this it is necessary that all subjects in every experiment should experience more or less the whole range of performance levels: that is, there should be at least one very good circuit condition, one of near average performance, and one very poor one, while the rest should not all cluster too closely about the same mean opinion score value. If one cannot be confident of this beforehand, it is advisable to carry out first a short informal test on the proposed set of circuit conditions, in order to find out whether the range is in fact covered; if not, the selection of conditions should be modified accordingly, otherwise the subjects' opinion scale will be distorted (the ''enhancement'' effect). Extra circuit conditions, not in themselves of direct interest to the experimenter, may be added to bring up the number to 9, 11 or 13, and to balance the range of performance more effectively.

Subjects generally expect to experience circuit conditions with various values of overall loss or sensitivity, which of course has a very strong influence on performance, and can be varied to provide the required range of circuit conditions. There are also important interactions between overall sensitivity and many other degradations. It is therefore highly desirable, even if overall sensitivity and its interactions are not the main objects of the investigation, to include some conditions differing from each other only in overall sensitivity.

If the investigation cannot be confined to 15 conditions, it is then spread over several experiments, each concentrating on a well defined part of the inquiry but overlapping the others so as to provide common ground.

3.4.1.3 *Eligibility of subjects*

Subjects taking part in the conversation tests are chosen at random from the Research Centre personnel, with the provisons that:

a) they have not been directly involved in work connected with assessment of the performance of telephone circuits; and

b) they have not participated in any subjective test whatever for at least the previous six months, and not in a conversation test for at least one year.

No steps are taken to balance the numbers of male and female subjects unless the design of the experiment requires it. Subjects are arbitrarily paired in the experimental design prior to the test and remain thus paired for its duration.

3.4.1.4 *Environment*

Subjects are seated in separate sound-proof cabinets near the point from which the experiment is controlled. Room noise is fed in with the required spectrum (usually the Hoth spectrum) at the required level (usually 50 dBA), measured with a Bruel and Kjaer Precision Sound Level Meter type 2206, used with the '' *A* weighting ´characteristic.If different conversations in the same experiment require different room noise levels, then care is taken to prevent the transitions from being too obvious to the subjects: ideally, room noise should be changed only when subjects are out of the sound-proof rooms.

3.4.1.5 *Methods of establishing the connection*

The telephone sets used by the subjects are normal in appearance and feel — usually identical to the standard British Telecom Telephone No. 706, unless the experiment specifically concerns handsets of other types. The means of establishing telephone contact between subjects is made as realistic as possible. The calling subject, on lifting the handset, obtains dialling tone, and has to dial or key a prescribed number to obtain the connection. Ringing tone occurs after a suitable fixed delay, and the other party's bell or tone-caller is operated after a further fixed delay. Wrong numbers are rewarded by the ''Number Unobtainable'' tone.

3.4.1.6 *Conversation task*

Every effort is made to ensure that conversations are purposeful, and that subjects have full opportunity to exploit the transmission capabilities of the test circuit. A task involving sorting pictures into an order of merit has been found suitable for this purpose and sufficiently interesting to the subjects. The pictures, covering a wide variety of topics, are samples of the standard postcard-sized illustrations offered for sale at several different museums, art galleries and similar institutions. These cards are individually numbered on the back, and assembled arbitrarily into sets of six cards each, every set having an exact duplicate.

The subject is instructed to consider these pictures for display in a public place, and, before each conversation, to arrange the cards of a particular set in his personal order of preference for this purpose; the other subject does the same with his copy of the same set. When contact is established via the test circuit, the subjects have to negotiate an agreed

order of preference and write this down at the end of the conversation. The duration of each conversation is thus determined by the subjects themselves. Occasionally a conversation may be very long because both subjects are intensely interested in the pictures, or — as happens in less than 1% of cases — very short because both have independently chosen the same order of preference and have little to discuss, but even in these cases it is highly desirable to allow the subjects to decide for themselves how long to converse. After the end of the conversation they express independent opinions of the connection by marking a form provided: one version of this form is reproduced in Annex A.

Some variations of the task (such as numbering the same pictures differently for the two subjects) have been explored, but none has been definitively adopted. No other type of task has been found to have any greater advantages for the purpose, though several types have been tried.

3.4.1.7 *Preparations for an* n × n *experiment*

From a list of all subjects available, the experimenter randomly chooses a sufficient number of those eligible according to the criteria given in § 3.4.1.3 above. He contacts these by telephone to ask whether they are willing to participate at certain times, which have to be arranged in such a way that subjects who converse together on their first visit remain paired for their subsequent visits in the same experiment. A standard letter is sent to each subject, confirming the time and place of each appointment, and explaining in some detail what will be required of the subjects in the experiment: the text of a typical letter is reproduced in Annex B.

The experimenter prepares schedules, based on the experimental design, showing in what order conditions must be administered to each pair of subjects, with which picture sets, which party initiates the call in each case, and any other necessary details. Space is left for filling in information that

becomes available as the experiment proceeds: consecutive conversation number, duration of conversation, identity of tape reel used for recording, comments about faults or unusual events, and so on. Opinion forms (Annex A) are also prepared for each conversation. However, in order to avoid duplicating or altering too many entries, some items are not filled in until they are certain: for example, the actual names of the subjects are liable to change until they actually arrive for their first visit.

Both in the letter and in any discussions with the subjects, great care is taken not to communicate to the subjects any knowledge about the nature of the circuit conditions. The opinion forms do not even carry any number or code identifying the circuit condition — this information is obtained from the schedule and added to the forms after they have been collected from the subjects.

3.4.1.8 *Procedure*

When subjects arrive for their first visit, they are asked whether they have read and understood the letter. Any obscurities are clarified, and opportunity is given for asking questions. The sound-proof rooms and their facilities are demonstrated. Subjects are informed how many calls will be comprised in this visit. Forms are handed to the subjects, and they are then left to prepare for the preliminary conversation. On subsequent visits the subjects are merely informed that the procedure will be the same as before, with possibly a different number of calls.

At the beginning of each conversation, the subjects take out the specified picture set from a box on the desk, arrange the pictures in order of preference, and fill in the appropriate part of the opinion form. When both subjects have done this, the experimenter gives one of them the signal to initiate the call. The subjects are then completely free to determine the course of the conversation, except that they must not discuss their opinions of the connection. When they have written down their agreed order of preference for the pictures, terminated the conversation, and recorded their scores

(Excellent, Good, Fair, Poor or Bad) and their answer to the ‘‘Difficulty’’ question (Yes or No), the experimenter contacts each in turn by telephone to ask what answer he has given to the ‘‘Difficulty’’ question; if the answer is ‘‘Yes’’, the experimenter asks the subject to explain briefly (in his own words) the nature of the difficulty. The reply is noted, but neither the subject nor the experimenter is expected to attempt precise formulations: it is essential not to prompt the subjects, and in any case the classification of difficulty has been found far less useful than the undifferentiated percentage ‘‘Difficulty’’ itself.

After this the experimenter requests the subject to put away the form in an envelope provided, and then tells him to start sorting out the next set of pictures, or, as the case may be, to wait to be released from the sound-proof room.

Both the conversations between subjects and the conversations between experimenter and subject are tape-recorded.

3.4.1.9 *Treatment of results*

The results from each conversation comprise two opinions on the scale Excellent-Good-Fair-Poor-Bad (scored respectively 4, 3, 2, 1, 0), two votes on the Difficulty scale (scored 1 = Yes, 0 = No), two speech levels (measured from tape recordings) and one value of duration. In particular cases information may be collected about other variables also; for example, video recordings may be made in order to observe how subjects hold their handsets.

Analysis of variance is applied separately to each variate (opinion score, speech level, etc.) in order to test the significance of circuit-condition features and other effects, and to find confidence intervals for the means. With a binary variate like ‘‘Difficulty’’ this process must be regarded with some reservations. There is usually less scope for curve-fitting than in listening experiments, simply because there are far fewer pairs of coordinate values available.

3.5 *Quantal-response detectability tests*

The best method for obtaining information on the detectability of some analogous property of a sound (such as echo), as a function of some objective quantity (such as listening level), is a quantal-response method similar in principle to that mentioned in § 3.1 above for loudness balancing. The main difference is that the subject’s response is not a decision in the form ‘‘Reference’’ or ‘‘Test’’ (the designation of the louder of two circuits), but a vote on a scale such as:

Opinion scale 6A

A Objectionable

B Detectable

C Not detectable

where B is understood to mean ‘‘Detectable but not objectionable’’.

Scales of this sort, usually with three points, may be used in a variety of quantal-response tests; for example the scale as shown above may be used where the stimulus is echo, reverberation, sidetone, voice-switching mutilation, or interfering tones, while crosstalk and perhaps echo in some circumstances may be judged on the scale Intelligible — Detectable — Not detectable.

It is sometimes permissible to regard these votes as opinion scores, with values 2, 1, 0 respectively, and treat them in the same sort of way as one would treat listening or conversation opinion scores. But this is often unsatisfactory because the decisions on such a scale as 6A are not really equivalents of responses on a continuous scale — as votes on such scales as 4A may be legitimately taken to be — but effectively embody two distinct dichotomies (for example detectable/not detectable and objectionable/not objectionable), which though not independent may nevertheless call different

psychological processes into action: in other words, Objectionability or Intelligibility differs in kind, not merely in degree, from Detectability, and often has a different standard deviation. For this reason a more profitable method of analysis is to express the probability of response according to each dichotomy separately, as a function of some objective variable, by fitting probit or logit equations, and then using the quantiles or other parameters as a basis of comparison between circuit conditions, in a manner analogous to that used in applying articulations scores.

The actual conduct of experiments of this type resembles that of listening-effort tests (see § 3.3.1 above), but there are some differences. In particular it is advisable that the first presentation of the signal in each run should be at a high listening level, so that the listener is left in no doubt what kind of signal is a candidate for his decisions. Where sidetone or echo is involved, the subject will be required to talk as well as listen.

Simple audiometric measurements, as described in Recommendation P.78, are usually performed on subjects who participate in these experiments, so that results can be expressed relative to their threshold of hearing.

For examples of the application of these techniques, see References [4] and [5].

Noise and other disturbances are sometimes investigated by means of responses on a scale with many more points; for example, Opinion scale 5 with seven points ranging from ‘‘Inaudible’’ to ‘‘Intolerable’’. These scales are more nearly of the quantized-continuum type, like Opinion scale 4A, and can be treated similarly.

4 Recommendations and other CCITT studies relying on Methods a) to e) of § 2 above:

a) Many Recommendations include requirements based originally on reference equivalents, later on corrected reference equivalents, and more recently on loudness ratings of which Recommendations P.12 (*Orange Book*), G.101 [6], G.103 [7], G.111 [8], G.120 [9] and G.121 [10] are examples.

b) Recommendation P.12 used to require certain articulation values to be satisfied but the method is now mainly used for diagnostic purposes. See Recommendation P.45.

c) Various Questions, for example Question 4/XII [11], Question 14/XII [12] and Supplement No. 3 at the end of this fascicle.

d) Various Questions, for example Question 4/XII [11], Question 9/XII [13], Ques tion 14/XII [12] and Supplement No. 3 at the end of this fascicle.

e) Various Questions, for example Question 9/XII [13] and References [14], [15] and [16].

5 General comments on subjective methods used in the laboratory

More detailed information on the conduct of subjective tests and interpretation of their results are given in Recommendation P.74 and Reference [2]. A rather broad survey of the relationship between various methods is given in Reference [17].

When used to provide information to assist in transmission planning of telephone networks, subjective methods should be employed with the following considerations in mind:

a) A clear description must be available of the type of telephone connections to which the results are to be applied. This is provided by formulating appropriate hypothetical reference connections (HRCs) (see Recommendation G.103 [7]).

b) The levels, transmission losses, sending and receiving reference equivalents, etc., of the HRCs must guide the establishment of laboratory arrangements and the conduct of the tests. Speech spectra and levels must be properly chosen to correspond to those at the various points in the HRC.

c) Subjects must be drawn from an appropriate population. For example, if audiograms are obtained from subjects participating in a conversation experiment, this information should not be used to reject any subjects, because the resultant bias in the sample would make the conclusions applicable only to users with a certain range of hearing sensitivity. For this reason it is safest to collect auxiliary information of this type only after the subjects have finished their main task.

d) Subjects must be treated within the experiments so that the results obtained are valid for the desired applications. This is the reason for taking the precautions described above (§ 3.3.1) to ensure that subjects’ judgements are not distorted by the range of conditions and levels chosen, or by the order of presentation; and to make the procedure in conversation tests (§§ 3.4.1.5 to 3.4.1.7) natural yet standardized.

e) Suitable experimental designs must be used so that the results can be properly analyzed and confidence intervals estimated.

f) Uncontrolled variation in some feature of the transmission path is sometimes unavoidable: for example the requirement may be to conduct a listening test over a fading radio link, or a conversation test over a TASI link with freeze-out determined by real traffic. In such cases it is advisable to collect not only the subjects' responses but also contemporary information on the values of the related fluctuating quantities: signal strength in the first case, freeze-out fraction or number of channels occupied in the second. The technique known as analysis of covariance (Reference [18]) is the appropriate method for processing this information on concomitant variables, as they are called, in conjunction with the responses (main variables).

g) Even with proper precautions under c), d), e) and f), reliance should not be placed on absolute values of scores unless ‘‘control’’ conditions (e.g. a set of reference conditions) are included within the experiment. However, relativities between scores obtained from different circuit conditions within the same experiment are more reliable.

h) A set of reference conditions will make it possible to express results as ratings in terms of equivalent settings of some reference device — attenuator, noise source, modulated noise reference unit (see Recommendation P.81), etc. This enables much more reliable comparisons to be made with information from other sources.

i) Results of subjective experiments should always be reviewed for internal consistency and compared with expected results (derived from previous experience or from a theoretical model) before being applied.

6 Objective methods

Clearly the ultimate aim must be to attain the capability of assessing telephony transmission performance purely in terms of the objective characteristics of the telephone connections concerned. This aim is partly satisfied by use of tabulated information based on previous laboratory and other tests: an example of such usage appears in Reference [19]. Considerable progress has now been made towards the prediction of assessment scores, speech levels, etc. by use of subjective modelling as described in Supplement No. 3 at the end of this fascicle and Reference [20]. British Telecom is now updating its tabulated information using this method.

The modelling technique makes it possible to treat many other important features like attenuation/frequency distortion and sidetone in a much more general manner. For example, by making due allowance for the part played by high sidetone level — which is a very potent degradation in connections of poor transmission performance — it makes clear why sensible limits for overall loss and noise cannot be fixed without regard to sidetone suppression.

ANNEX A

(to Supplement No. 2)

Opinion form 12A

Test

Name Cabinet No.

1 Before starting your call, please take out picture set and arrange the cards in order of preference. Record this order in the boxes below, using the numbers on the backs of the cards for identification.

H.T. [T1.2]

lw(48p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) , ^ | l | l | l | l | l | l. lw(48p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) , ^ | l | l | l | l | l | l. Your Order of Preference 1st 2nd 3rd 4th 5th 6th

_

Table A-1 [T1.2], p.

2 If you receive a green GO AHEAD signal, then call your partner on

3 You may enter your partner's picture-card order here if you find this helps you in the discussion.

H.T. [T2.2]

lw(48p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) , ^ | l | l | l | l | l | l | l.
lw(48p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) , ^ | l | l | l | l | l | l | l. Partner's
Order of Preference 1st 2nd 3rd 4th 5th 6th
_

Table A-2 [T2.2], p.

4 When you have arrived at an agreed order of preference, please enter it here.

H.T. [T3.2]

lw(48p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) | lw(30p) , ^ | l | l | l | l | l | l | l.
lw(48p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) | cw(30p) , ^ | l | l | l | l | l | l | l. Agreed
Order of Preference 1st 2nd 3rd 4th 5th 6th
_

Table A-3 [T3.2], p.

Then replace your handset.

5 Please mark, with a cross, your opinion of the telephone connection you have just been using. *N.B.* — Please do not discuss your opinion with your partner.

H.T. [T4.2]

Excellent	Good	Fair	Poor	Bad

Table A-4 [T4.2], p.

6 Did you or your partner have any difficulty in talking or hearing over the connection?

H.T. [T5.2]

YES	
NO	

Table A-5 [T5.2], p.

If the answer is YES, please explain briefly what the difficulty was when the operator contacts you again.

H.T. [T6.2]

lw(60p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) . lw(60p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) | lw(12p) . FOR R13.4 USE
_

Tableau A-6 [T6.2], p.

ANNEX B

(to Supplement No. 2)

(Standard letter sent to subjects)

Name

Group

R13.4 SUBJECTIVE TEST No.

Thank you for agreeing to take part in this experiment.

As arranged earlier by telephone, we should like you to come to Room , Floor 3, Main Laboratory Block, at the following times.

Time Day Date

On arrival, ask for , quoting the above subjective test number. You will be reminded by telephone shortly before each visit is due. You may book your time to project appointment, or if you need further information, contact on Ipswich 64

The experiment to which you have been invited forms part of a series concerned with the transmission performance of telephone connections. You will be asked to converse with another volunteer over particular telephone connections, and it is hoped that the tasks we shall give you will lead to vigorous conversations devoted to discussion and negotiation.

In the test room you will be provided with a set of six picture cards. You are asked to imagine that you and your partner are responsible for choosing some of these (enlarged if necessary) to be displayed in a public place such as the Staff Restaurant — either as items of general interest or simply as decoration. Before each call you should arrange all six cards in your order of preference, and write the six identification numbers in this order on the form provided. Your partner, in another room, will have an identical set of pictures, and his order of preference will probably be different from yours. One of you will then be requested to make a telephone call to the other. The aim of the ensuing conversation will be to negotiate with your partner so as to arrive at a compromise order which satisfies you both. At the end of the conversation you should replace the handset and enter the six numbers in the finally agreed order on the form. You must also mark the appropriate box to indicate your opinion of the connection. After this the operator will contact you and tell you what to do next. Subsequent conversations will be similar, but with different sets of pictures.

In the whole experiment there will be a total of calls spread over the visit(s) arranged as above. Full instructions will be given when you arrive. Please bring this letter with you, and also your glasses if you normally wear any.

Thank you once again for your co-operation.

(date)

(date)

References

[1] CCITT — Question 15/XII, Contribution COM XII-No. 1, Study Period 1981-1984, Geneva, 1981.

[2] RICHARDS (D. | .): Telecommunication by speech, *Butterworths* , London, 1973.

[3] SULLIVAN (J. | .): Is transmission satisfactory? Telephone customers help us decide, *Bell Labs Record* , pp. 90-98, March 1974.

[4] RICHARDS (D. | .): Telecommunications by speech, § 3.5.3. *Butterworths* , London, 1973.

[5] *Ibid.* , § 4.5.1.

[6] CCITT Recommendation *The transmission plan* , Vol. III, Rec. G.101.

[7] CCITT Recommendation *Hypothetical reference connections* , Vol. III, Rec. G.103.

[8] CCITT Recommendation *Loudness ratings (LRs) in an international connection* , Vol. III, Rec. G.111.

[9] CCITT Recommendation *Transmission characteristics of national networks* , Vol. III, Rec. G.120.

[10] CCITT Recommendation *Loudness ratings (LRs) of national systems* , Vol. III, Rec. G.121.

[11] CCITT — Question 4/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[12] CCITT — Question 14/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[13] CCITT — Question 9/XII, Contribution COM XII-No. 1, Study Period 1981-1984, Geneva, 1981.

[14] RICHARDS (D. | .) and BUCK (G. | .): Telephone echo tests, *P.I.E.E.* , 107B, pp. 553-556, 1960.

[15] CCITT — Contribution COM XII-No. 171, Study Period 1977-1980, Geneva, 1979.

[16] CCITT — Contribution COM XII-No. 132, Study Period 1977-1980, Geneva, 1979.

[17] CCITT — Question 7/XII, Annex 1, Contribution COM XII-No. 1, Study Period 1981-1984, Geneva, 1981.

[18] SNEDECOR (G. | .) and COCHRAN (W. | .): Statistical methods, Chapter 14, 6th edition, *Iowa State University Press* , 1967.

[19] CCITT — Contribution COM XII-No. 173, Study Period 1977-1980, Geneva, 1979.

[20] RICHARDS (D. | .): Calculation of opinion scores for telephone connections, *Proc. IEE* , 121, pp. 313-323, 1974.

Bibliography

BRAUN (K.): Die Bezugsd"ampfung und ihre Berechnung aus der Restd"ampfungskurve (Frequenzkurve) eines "Ubertragungssystems; *T.F.T.* , Vol. 28, pp. 311-318, August 1939.

BRAUN (K.): Theoretische und experimentelle Untersuchung der Bezugsd"ampfung und der Lautst"arke; *T.F.T.* , Vol. 29, pp. 31-37, No. 2, 1940.

BLYE (P. | .), COOLIDGE (O. | .) and HUNTLEY (H. | .): A revised telephone transmission rating plan; *B.S.T.J.* , Vol. 34, pp. 453-472, May 1955 (reproduced in the *Red Book* , Vol. I, pp. 636-651, ITU, Geneva, 1957, and Vol. V, pp. 607-624, ITU, Geneva, 1962).

BRAUN (K.): Image attenuations of microphone and receiver insets; *N.T.Z.* , No. 8, pp. 365-370, 1960 (translated in the *Red Book* , Vol. V | fIbis , pp. 255-265, ITU, Geneva, 1965).

FRENCH (N. | .) and STEINBERG (J. | .): Factors governing the intelligibility of speech sounds; *J.A.S.A.* , Vol. 19, p. 89, Jan. 1947.

RICHARDS (D. | .) and ARCHBOLD (R. | .): A development of the Collard principle of articulation calculation; *P.I.E.E.* , Vol. 103, Part B, Sept. 1956 (*Red Book* , Vol. I, Question 7 of Study Group 12, Annex 4, ITU, Geneva, 1956).

Contribution by the Italian Administration to the study of objective methods for measuring reference equivalent and articulation reference equivalent, *Red Book* , Vol. I, Question 7 of Study Group 12, Annex 3, ITU, Geneva, 1956.

FLETCHER (H.) and GALT (R. | .): The perception of speech and its relation to telephony; *J.A.S.A.* , Vol. 22, p. 89, March 1950 (reproduced in the following work, Chapters 15-17).

FLETCHER (H.): Speech and hearing in communication, *D. Van Nostrand* , New York, 1953.

Tonality method studied by the U.S.S.R. Administration to determine articulation; *Red Book* , Vol. V, Part II, Annex 31, ITU, Geneva, 1962.

Method used by the Swiss Telephone Administration for the determination of transmission quality based on objective measurements; *Red Book* , Vol. V, Part II, Annex 30, ITU, Geneva, 1962.

LALOU (J.): Calculation of telephone transmission performance by information theory, *Red Book* , Vol. V | fIbis , Question 7/XII, Annex 2, ITU, Geneva, 1965.

SIVIAN (L. | .): Speech power and its measurement, *B.S.T.J.* , 8, pp. 646-661, 1929.

LOYE (D. | .) and MORGAN (K. | .): Sound picture recording and reproducing characteristics, *J. Soc. Motion Picture Engineers* , 32, pp. 631-647, 1939.

RICHARDS (D. | .): Some aspects of the behaviour of telephone users as affected by the physical properties of the circuit. Communication Theory, *Butterworths Scientific Publications* , pp. 442-449, 1953.

ZAITSEV (T. | .): Correlation method for determining the fidelity and intelligibility of speech transmitted over telecommunication channels, *Elektrosvyaz* , 10, pp. 38-46, 1958.

LICKLIDER (J. | . | .), BISBERG (A.) and SCHWARZLANDER (H.): An electronic device to measure the intelligibility of speech, *Proc. Nat. Electronics Conf.* , 15, pp. 329-334, 1959.

RICHARDS (D. | .) and SWAFFIELD (J.): Assessment of speech communication links, *P.I.E.E.* , 106B, pp. 77-89, 1959.

RICHARDS (D. | .): Conversation performance of speech links subject to long propagation times, International Conference on Satellite Communication, *Inst. Elec. Engrs.* , pp. 247-251, London, 1962.

RICHARDS (D. | .): Transmission performance of telephone connections having long propagation times, *Het PTT-Bedrijf* , 15, pp. 12-24, 1967.

BOERYD, (A.): Subscriber reaction due to unbalanced transmission levels, *ibid* , pp. 39-43.

RICHARDS, (D. | .): Distortion of speech by quantizing, *Electronics Letters* , 3, pp. 230-231, 1967.

GOLDMAN-EISLER (F.): Sequential temporal patterns and cognitive processes in speech, *Language and Speech* , 10, pp. 122-132, 1967.

Supplement No. 3

MODELS FOR PREDICTING TRANSMISSION QUALITY FROM OBJECTIVE MEASUREMENTS

Models for predicting the subjective opinion of telephone connections, using data from objective measurements, are currently under study in Question 7/XII. It has not been possible up to now to recommend a single model applicable over a wide range of transmission impairments, but the methods described in §§ 1, 2, 3, 4 below have been proposed by several Administrations.

1 Transmission rating models *(Geneva, 1980; modified at Malaga-Torremolinos, 1984) (Quoted in § 3 of Recommendation P.11)*

(Contribution by the Bell Communications Research, Inc.)

1.1 *Introduction*

This Section describes transmission rating models which can be used to estimate the subjective reaction of telephone customers to the transmission impairments of circuit noise, overall loudness rating, talker echo, listener echo, attenuation distortion (including bandwidth), quantizing distortion, room noise and sidetone.

The models for circuit noise overall loudness rating (OLR) and talker echo are based on several conversational tests conducted at Bell Laboratories in the period from 1965 to 1972 to evaluate the subjective assessment of transmission quality as a function of circuit noise, overall loudness rating, talker echo path loss and talker echo path delay [1]. These tests involved several hundred subjects

This Section (former Supplement No. 3, *Red Book*), reflects in part work performed at AT&T Bell Laboratories prior to 1 January 1984.

and several thousand test calls, Several tests were conducted on normal business calls. Others were conducted in the laboratory. All of the tests employed a 5-category rating scale: excellent, good, fair, poor and unsatisfactory.

The essential features of the models were originally derived in terms of loudness loss of an overall connection *in dB* (as measured by the Electro-Acoustic Rating System, *EARS*) and circuit noise *in dBmp* at the input to a reference receiving system (electric-to-acoustic efficiency as measured by the *EARS*) [2]. The effects of talker echo were later incorporated in terms of loudness loss of the echo path *in dB* (as measured by the *EARS*) and round trip delay of the echo path *in milliseconds* . Experimentally determined correction factors were used to convert the models to loudness ratings according to Recommendation P.79.

The original model for listener echo was based on a series of four listening-type subjective tests conducted at Bell Laboratories in 1977 and 1978 [4]. Subsequent test results led to an alternative form of the model [5], [6]. The subjective tests included conditions in which the listener echo path loss was flat or frequency-shaped by selective filtering. A weighted echo path loss is defined to provide a weighting of the frequency-shaped test conditions so that subjectively equivalent test conditions have the same transmission rating.

The model for quantizing distortion is based on a series of five subjective tests conducted to evaluate the performance of various digital codec algorithms [7], [8], [9].

The model for bandwidth and attenuation distortion is based on tests conducted in 1978 [10].

The model for room noise is based on unpublished tests conducted in 1976. Opinion ratings of transmission quality on a five-category scale were made by 40 subjects for 156 conditions having various combinations of room noise, speech level, circuit noise and sidetone path loss. The samples of room noise were presented from tape recordings made in an airlines reservations office. A model was fitted to the test results in terms of the circuit noise which produced the same quality ratings as given levels of room noise.

The model for sidetone is based on tests conducted in 1980 [11].

All of the tests were conducted with Western Electric 500-type telephone sets or equivalent. The procedures used in the analysis of the subjective tests results and the derivation of the transmission rating scale are outlined in Reference [1]. 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.

The models incorporate the concept of a transmission rating scale recognition that subjective test results can be affected by various factors such as the subject group, the type of test, and the range of conditions which are included in the test. These factors have been found to cause changes in both the mean opinion score of a given condition and in the standard deviation. Thus, there are difficulties in trying to establish a unique relationship

between a given transmission condition and subjective opinion in terms of mean opinion score or percent of ratings which are good or excellent. The introduction of a transmission rating scale tends to reduce this difficulty by separating the relationship between transmission characteristics and opinion ratings into two parts. The first part, the transmission rating as a function of the transmission characteristic, is anchored at two points and tends to be much less dependent on individual tests. The second part, the relationship between the transmission rating and subjective opinion ratings, can then be displayed for each individual test.

The transmission rating scale for overall loudness rating and circuit noise was derived such that it is anchored at two points as shown in Table 1-1.

H.T. [T1.3]

TABLE 1-1

Overall loudness rating (dB) Circuit noise (dBmp) | ua) }	{ Transmission rating	
16	—61	80
31	—76	40

a) The circuit noise values are referred to a receiving system with a receiving loudness rating (RLR) = 0 dB.

Table 1-1 [T1.3],

These anchor points were selected to be well separated but within the range of conditions which are likely to be included in a test. The rating values are such that most connections will have positive ratings between 40 and 100. Transmission ratings for other combinations of loudness rating and circuit noise are relative to those for these two anchor points.

This Section presents the transmission rating models in terms of overall loudness rating of an overall connection in dB, circuit noise in dBmp referred to the input of a receiving system with a receiving loudness rating (RLR) = 0 dB, loudness rating of the talker echo path in dB, and round-trip delay of the talker echo path in milliseconds. Annex A illustrates representative opinion results.

1.2 *Transmission rating models*

1.2.1 *Overall loudness rating and circuit noise*

The transmission rating model for overall loudness rating and circuit noise is

$$R_{LN} = -26.76 - 2.257\sqrt{(L'_e - 8.2)^2 + 1} - 2.0294\, N'_F + 1.751\, L'_e + 0.02037\, L'_e N'_F \qquad (1\text{-}1)$$

where

L'_e is the OLR of an overall telephone connection (in dB).

N'_F is the total effective noise (in dBmp) referred to a receiving system with a 0 dB RLR. The total effective noise is obtained by the power addition of the circuit noise, N'_c, the circuit noise equivalent, N'_{Re}, of the room noise and the circuit noise equivalent, N'_{Qe}, of the quantizing noise.

N'_c is the circuit noise (in dBmp) referred to a receiving system with a 0 dB RLR.

N'_{Re} is the circuit noise equivalent (in dBmp) of the room noise referred to a receiving system with a 0 dB RLR. (See § 1.2.2.)

N'_{Qe} is the circuit noise equivalent (in dBmp) of the quantizing noise referred to receiving system with a 0 dB RLR. (See § 1.2.3.)

Transmission rating as a function of the OLR and circuit noise is shown in Figure 1-1. This figure uses a value of $N'_{Re} = -58.63$ dBmp. Bandwidth factor, k_{BW}, defined in § 1.2.4 is equal to unity.

1.2.2 *Circuit noise equivalent of the room noise*

The transmission rating model for the circuit noise equivalent, N'_{Re} (in dBmp), of the room noise is

N'

$$N'_{Rc} = N_R - 121 + 0.0078\,(N_R - 35)^2 + 10 \log_{10}\left[1 + 10^{-L'_s/10}\right] \qquad (1\text{-}2)$$

where

N_R is the room noise in dB(A) at the listening end

L'_s is the sidetone masking rating (in dB) of the listening end telephone set sidetone path

The circuit noise equivalent, N'_{Rc}, is plotted as a function of room noise in Figure 1-2.

Note — The transmission rating model for loudness rating and circuit noise is normally used with

$$N'_{Rc} = -58.63 \text{ dBmp.} \qquad (1\text{-}3)$$

This value was determined from analysis of the conversational tests results from which the transmission rating model for the overall loudness rating and circuit noise was originally formulated.

Figure 1-1, p.

Figure 1-2, p.

1.2.3 *Circuit noise equivalent of quantizing noise*

The transmission rating model for the circuit noise equivalent $N'_{Q\backslash de}$ (in dBmp) of quantizing noise is

$$N'_{Q\backslash de} = V - 2 - \mathrm{SNR} \tag{1-4}$$

where

V is the active speech level (in dBM) referred to a receiving system with a 0 dB RLR,

and

SNR is the signal-to-circuit noise ratio (in dB) which is judged to provide speech quality equivalent to the speech-to-speech correlated noise ratio, Q (in dB), as determined by a Modulated Noise Reference Unit (see Recommendation P.81).

SNR can be approximated by

$$\mathrm{SNR} = 2.36\,Q - 8 \tag{1-5}$$

from which

$$N'_{Q\backslash de} = V - 2.36\,Q + 6 \tag{1-6}$$

Based on a 1975-1976 Speech Level Survey, [12] the speech level for domestic North American connections can be approximated by

$$V = -9 - L'_e$$

from which

$$N'_{Q\backslash de} = -3 - L'_e - 2.36\,Q \tag{1-7}$$

Estimates of Q | or single codec pairs are given below for Pulse Code Modulation (PCM), Nearly-Instantaneous Companded modulation (NIC), Adaptive Differential Pulse Code Modulation (ADPCM) and Adaptive Delta Modulation (ADM). They apply to the particular algorithms described in References [7] and [9].

PCM: $Q = 0.78\,L - 12.9$ (1-8)

NIC: $Q = 0.74\,L - 2.8$ (1-9)

ADM: $Q = 0.42\,L + 8.6$ (1-10)

ADPCM: $Q = 0.98\,L - 5.3$ (1-11)

ADPCM-V: $Q = 1.04\,L - 4.6$ (1-12)

where

L is the line bit rate in kbit/s.

Note — The ADPCM algorithm with fixed predictor is described in Reference [13]. The ADPCM-V algorithm with adaptive predictor is described in Reference [9].

For connections with tandem codec pairs, the total Q | an be estimated as follows:

$$Q = -15 \log_{10}$$

$$\left[\sum_{i=1}^{f\,In} 10^{\,(em\ \frac{f\,IQ}{5}\, i\ f\,R} \right]$$

(1-13)

1.2.4 *Bandwidth and attenuation distortion*

The transmission rating model for overall loudness rating and circuit noise can be modified to include the effects of bandwidth (and attenuation distortion). The transmission rating, R_{LNBW} .PS 10 , for overall loudness rating, circuit noise and bandwidth is

$$R_{LNBW} \text{ .PS 10} = (R_{L\backslash dN} - 22.8)\ k_{B\backslash dW} + 22.8$$

(1-14)

where

$$k_{B\backslash dW} = k_1 k_2 k_3 k_4 \quad (1\text{-}15)$$

.sp 1

with

$$k_1 = 1 - 0.00148\,(F_l - 310) \quad (1\text{-}16)$$
$$k_2 = 1 + 0.000429\,(F_u - 3200) \quad (1\text{-}17)$$
$$k_3 = 1 + 0.0372\,(S_l - 2) + 0.00215\,(S_l - 2)^2 \quad (1\text{-}18)$$
$$k_4 = 1 + 0.0119\,(S_u - 3) - 0.000532\,(S_u - 3)^2 - 0.00336\,(S_u - 3)\,(S_l - 2) \quad (1\text{-}19)$$

and

F_l, F_u is the lower and upper band limits (in Hz) at which the acoustic-to-acoustic response is 10 dB lower than the response at 1000 Hz. (For F_u > 3200 Hz, a value of 3200 Hz should be used.)

S_l, S_u is the lower and upper inband response slopes (in dB/octave) below and above 1000 Hz, respectively, which would have the same loudness loss as the actual response shapes.

Figures 1-3 and 1-4 illustrate the effect of the band limits, F_l and F_u, and inband slopes, S_l and S_u, on the bandwidth factor, $k_{B\backslash dW}$.

Note — The functions for the bandwidth factor, $k_{B\backslash dW}$, have been selected such that $k_{B\backslash dW} = 1$ when F_l = 310 Hz, F_u = 3200 Hz, S_l = 2 dB/octave and S_u = 3 dB/octave. These response characteristics are rep resentative of those used in the tests to formulate the transmission rating model for overall loudness rating and circuit noise.

Figure 1-3, p.

Figure 1-4, p.

1.2.5 *Listener echo*

The transmission rating model for listener echo is

$$R_{LE} = 9.3\,(WEPL + 7)\,(D_L - 0.4)^{-0.229}$$

(1-20)

where

WEPL is the Weighted Listener Echo Path Loss (in dB) and

$$WEPL = -20 \log_{10}$$

[Unable to Convert Formula]

$$\int_{200}^{400}$$

10

—

$f\,IEPL\,(f)$

0

d
f

(1-21)

$EPL(f)$ is the echo path loss (in dB) as a function of frequency in Hz.

D_L is the round-trip listener echo path delay in milliseconds.

Transmission rating, $R_{L\backslash dE}$, as a function of the weighted echo path loss and listener echo-path delay is shown in Figure 1-5.

Figure 1-5, p.

The transmission rating for listener echo, R_{LE}, can be combined with the transmission rating for overall loudness rating and circuit noise to give an overall transmission rating as follows:

R

$LNLE$

[Unable to Convert Formula]
[Unable to Convert Formula]

(1-22)

Figure 1-6 provides curves generated by means of the above relationship for transmission rating as a function of weighted listener echo path loss and listener echo path delay in a connection with an overall loudness rating of 16 dB and a circuit noise of —56 dBmp referred to a RLR of 0 dB.

Figure 1-6, p.

Note — The preceding material is based on the use of a specific set of test results and the listener echo model of Reference [4]. Subsequently, new test results were reported in References [5] and [6] which also described studies of the two sets of tests results to see if a single model could be recommended. In general, the agreement between the two sets of results was good. However, the newer results had lower opinion ratings at delays less than about 3 ms. A conservative approach was to revise the original model to provide lower ratings at low delays while retaining the more critical predictions at higher values of delay. The following equation (1-20a) provides a satisfactory replacement for equation (1-20) which accomplishes this goal.

$$R_{LE} = 10.5\,(WEPL + 7)\,(D_L + 1)^{-0.25} \qquad (1\text{-}20a)$$

Reference [6] also proposed that Weighted Echo Path Loss (WEPL) in the original model be replaced by Scaled Weighted Echo Path Loss (SWEPL). The proposal defined

$$WEPL = SM + SF$$

where

SM is the singing margin,

SF is the shape factor

and then defined

$$SWEPL = SM + SF \frac{\mathrm{f\,ISM}}{+ SM}$$

Hence, like *WEPL* ,

$SWEPL = SM$, if $SF = 0$.

Also,

$SWEPL = WEPL$, for $SM >> 1$.

The effect of the shape factor is reduced as *SM* | pproaches zero. Thus, the shape effect is cut in half when *SM* is equal to unity, and approaches zero as *SM* approaches zero. This avoids the possibility of a positive *SWEPL* when singing margin has become negative. Although the use of *SWEPL* instead of *WEPL* will cause little change in most practical situations with typical values of *SM* , the concept is attractive in forcing the singing margin to be specifically taken into account and is easily accomplished by replacing *WEPL* by *SWEPL* in equation (1-20a).

1.2.6 *Talker echo*

The transmission rating model for talker echo is

$$R_E = 92.73 - 53.45 \log_{10} \left[\$\$1o\ 1 + D\ \$\$2u \sqrt{1 + \left(\frac{f\,ID}{80} \right)}\ \$\$2x\ 2\$\$\ 2e \right] + 2.277 E \qquad (1\text{-}23)$$

where

E is the OLR (in dB) of the talker echo path

D is the round-trip talker echo path delay in milliseconds.

Transmission rating as a function of talker echo path loss and delay is shown in Figure 1-7 and has been derived to exclude the effects of circuit noise and OLR. Transformation of the talker echo test results, which included selected values of OLR and circuit noise, to the transmission rating scale, was accomplished using the $R_{L\backslash dN}$ model.

The transmission rating model for the combined effects of OLR, circuit noise, echo path loss and echo path delay is

R_{LNE}

[Unable to Convert Formula]
[Unable to Convert Formula]

(1-24)

Figure 1-8 shows curves generated by means of the above relationship for the transmission rating as a function of talker echo path loss and delay in a connection with an OLR of 16 dB and circuit noise of —56 dBmp.

1.2.7 *Sidetone*

The transmission rating model for OLR, total effective noise and talker echo can be modified to include the effects of sidetone. The transmission rating, $R_{L\backslash dN\backslash dS\backslash dT}$, for OLR, total effective noise and sidetone is

$$R_{L\backslash dN\backslash dS\backslash dT} = K_{S\backslash dT} R_{L\backslash dN} \qquad (1\text{-}25)$$

and for talker echo and sidetone is

$$R_{E\backslash dS\backslash dT} = R_E + 2.6(12 - SL) - 1.5(4.5 - SR)^2 + 3.38.$$

(1-26)

The sidetone factor, $K_{S\backslash dT}$, is calculated from

$$K_{S\backslash dT} = 1.021 - 0.002(SL - 15)^2 + 0.001(SR - 2)^2 (SL - 15).$$

(1-27)

Figura 1-7, p.

Figura 1-8, p.

SL | s the sidetone masking rating (in dB), *SR* | s the sidetone response (in dB/octave) below 1 kHz. (The sidetone response above 1 kHz is 1.5 times greater.)

Figure 1-9 shows curves obtained by determining $R_{L\dN\dS\dT}$ | and $R_{E\dS\dT}$, then substituting these values for $R_{L\dN}$and $R_E$respectively in equation (1-24).

1.3 *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. The proportion of comments Good plus Excellent (G + E) or Poor plus Unsatisfactory (P + U) can be computed from the following equations:

G + E =
[Unable to Convert Formula]
[Unable to Convert Formula]
e
[Unable to Convert Formula]

d
t

(1-28)

P + U =
[Unable to Convert Formula]
(*em* ∞
∫
B
e
[Unable to Convert Formula]

d
t

(1-29)

where A and B are given below for data bases of primary interest.

For each data base listed below, the relationship between the subjective judgements and transmission rating is shown in Figure 1-10.

Data Base	*A*	*B*
1965 Murray Hill SIBYL Test	(R-64.07)/17.57	(R-51.87)/17.57
CCITT Conversation Tests	(R-62)/15	(R-43)/15
Long Toll Interviews	(R-51.5)/15.71	(R-40.98)/15.71

Sidetone Response:	*Below 1 kHz*	*Above 1 kHz*
	0	0
	+3.0	+4.5
	+6.0	+9.0

The three data bases reflect different relationships between the transmission rating scale and opinion ratings as determined in different tests as indicated below: *1965 Murray Hill SIBYL Test* — Opinions on actual intra-building business calls. *CCITT Conversation Tests* — Composite model of opinion in laboratory conversation tests reported to the CCITT in the 1973-1976 Study Period (see [3]). *Long Toll Interviews* — Opinions expressed by North American Telephone customers when interviewed following a call on a long toll connection.

2 Prediction of transmission qualities from objective measurements *(Geneva, 1980; modified in Malaga- Torremolinos, 1984)* *(Quoted in Recommendation P.11)* (Contribution from British Telecom)

Summary

British Telecom makes extensive use of a theoretical model for predicting the transmission performance of telephone connections brief description is here given of the structure of this model, and of the computer program CATNAP, which embodies a simplified form of the model for routine use, together with facilities for specifying connections in a convenient practical way.

Formerly, Supplement No. 4, *Red Book* .

Figure 1-9, p. 16

Figure 1-10, p.17

2.1 *Types of model*

Question 7/XII [14] recognises two types of ‘‘model’’ for predicting the performance of complete telephone connections in conversation. The first kind, exemplified in Section 1 of this supplement, involves purely empirical treatment of basic observations, and might lead to a set of tables, graphs or relatively simple formulae, representing performance as a function of certain objective quantities. In a model of this type, where attention is focussed entirely on the correspondence between input (objective quantities) and output (subjective performance), the *form* of the functions employed has no significance in itself. For convenience, simplicity is usually sought, but is obtained at the expense of generality. Interactions between different degradations are often difficult enough to treat in any case; but, besides, a purely empirical model must usually be completely revised when a new degradation is brought in. For example, suppose relationships have been established between loss, noise and opinion score for one particular bandwidth: changing that bandwidth to a new constant value will necessitate a redetermination of the functions — not just a constant adjustment of the output. In short, it is unreasonable to expect that a purely empirical model could have more than limited success in predicting performance.

Models of the second type (mentioned in [15]) are intended to overcome these disadvantages by making the structure of the evaluation process reflect the cause-and-effect relationships which lead from the input (properties of the connection; acoustic environment; characteristics of the participants’ hearing, speech sounds and language systems, etc.) to the output (participants’ satisfaction or estimate of performance). Such a model is inherently more complicated, and requires more work to develop initially, but can then be extended and applied with much greater ease and confidence. Numerical parameters may and do require revision as more reliable data become available, but the structure, if well chosen, will only rarely require major alterations. As a research tool, such a model is much more powerful in its capability of generating hypotheses to be tested than a collection of useful but arbitrary formulae. As a planning or application tool, it lends itself easily to being embodied in a computer program, to which readily available data (such as losses and line lengths) can be supplied as input.

2.2 *Model and programs: SUBMOD, CATPASS and CATNAP*

The model here described is of the more fundamental type. It is intended to predict loudness judgements, listening-effort scores, conversation-opinion scores and vocal levels from objective information supplied. It is embodied in a program called SUBMOD (mnemonic for SUBJECTIVE MODEL) which accepts the overall frequency responses of the speech-transmission paths as input, and makes provision for changing the parameters of the model in order to improve agreement between theory and observation. Reference [16] describes an earlier version of the same model.

In its present state of development the model deals fairly successfully with the subjective effects of circuit loss, attenuation-frequency distortion, circuit noise, quantizing noise, room noise, and sidetone paths, for a reasonably wide range of values of these characteristics in any combination. Effects of some other phenomena can also be approximately estimated, but are not yet incorporated in the model. No attempt has yet been made to cater for features such as voice-switching effects, or vocoding and other sophisticated schemes for reducing information rate. Compare the groups of factors listed in Question 7/XII [14].

The program CATPASS [16] — a mnemonic for COMPUTER-AIDED TELEPHONY PERFORMANCE ASSESSMENT — incorporated the same model in a simplified, fixed-parameter implementation, together with facilities for calculating the sensitivity-frequency response of a complete connection formed by concatenating common pieces of apparatus such as telephones, cables, feeding bridges, junctions, and filters. It was similar to the system described in [17] and [18], but the program was differently organized. However, CATPASS could handle symmetrical connections only — that is, those for which transmission, room noise, sidetone and all other relevant features were the same for both participants. It was superseded by a program called CATNAP (COMPUTER-AIDED TELEPHONE NETWORK ASSESSMENT PROGRAM), which incorporated an extended form of the fixed-parameter model, allowing asymmetry in the connections, as well as containing facilities for assembling performance statistics on sets of connections. See [19].

CATNAP has been superseded in turn by CATNAP83, in which three main changes have been made:

a) minor improvements to the subjective model;

b) calculation of loudness ratings according to Recommendation P.79, instead of the provisional version P.XXE [20] which (notwithstanding the statement made in the earlier version of this Supplement [21]) was used for calculating loudness ratings in CATNAP;

c) introduction of more flexibility to allow parameters such as the earphone coupling loss factor (L_E) to depend on the particular type of handset.

2.3 *Situation to be represented*

Let A and B denote two ‘‘average’’ participants in a telephone conversation over a link terminated in handset telephones, located in rooms with no abnormal reverberation and with specified levels of room noise. ‘‘Average’’ is intended to convey that the participants have representative hearing and speaking characteristics and a normal attitude towards telephone facilities, so that their satisfaction with the telecommunication link may be measured by the mean Conversation Opinion Score (Y_C) and the Percentage Difficulty (%D) that would be obtained from a conversation test, as described in Supplement No. 2. Y_C can take any value between 4 and 0, the scale being: 4 = EXCELLENT, 3 = GOOD, 2 = FAIR, 1 = POOR, 0 = BAD. %D can of course take any value between 0 for the best connections and 100% for the worst.

For a given connection, the quantities of chief interest are Y_C, %D | and the speech level, for each participant. However, other useful auxiliary quantities are computed in the course of the evaluation, such as the loudness ratings of the various paths (calculated according to Recommendation P.79), and Y_{LE}, the mean Listening Effort Score that would result from a listening opinion test conducted as outlined in Supplement No. 2. In a listening test of this type, lists of sentences at a standard input speech level are transmitted over the connection and the listener expresses an opinion, at a number of different listening levels, on the ‘‘listening effort’’ according to the following scale:

Effort required to understand the meanings of sentences

A Complete relaxation possible; no effort required

B Attention necessary; no appreciable effort required

C Moderate effort required

D Considerable effort required

E No meaning understood with any feasible effort.

The votes are scored A = 4, B = 3, C = 2, D = 1, E = 0, and the mean taken over all listeners is called the Listening Effort Score, $Y_{L\backslash dE}$, for each particular listening level and each circuit condition.

More detailed information about both conversation tests and listening tests may be found in [22], and also in Supplement No. 2.

2.4 *Outline of the model*

The model requires the following inputs:

1) overall sensitivity-frequency characteristic of each transmission path (talker's mouth to listener's ear via the connection) and sidetone path (each talker's mouth to his own ear). These sensitivities may be either measured by the method described in Recommendation P.64 or calculated as explained in Reference [17];

2) noise spectrum and level at each listener's ear, composed of noise arising in the circuit, room noise reaching the listening ear direct, and room noise reaching the listening ear via the sidetone path. In the absence of specific measurements, standard noise spectra and levels are taken; e.g. room noise with Hoth spectrum at 50 dBA, circuit noise with bandlimited spectrum at a specified psophometrically weighted level;

3) average speech spectrum and average threshold of hearing, as given for example in [23].

From these data the loudness ratings are calculated. With speech level fixed, $Y_{L\backslash dE}$ | and a provisional value of Y_C are evaluated for each participant. The relationships between Y_C and speech level at each end are then used to refine the values of both, so that the final estimates represent performance at realistic conversational speech levels.

2.5 *Calculation of loudness and loudness ratings*

The model starts by setting the speech level emitted from each talker to a standard value and calculating the resultant spectrum and level of both speech and noise at each listener's ear. The loudness of received speech is calculated as a function of signal level, noise level and threshold of hearing, integrated over the frequency range extending normally from 179 to 4472 Hz (14 bands, the lowest centred at 200 Hz and the highest at 4000 Hz). The loudness of the sidetone speech is calculated similarly, but with an allowance for the additional masking effect of speech reaching the ear naturally (via the air path and the bone-conduction path). By comparison with the loudness of speech transmitted over an IRS (Intermediate Reference System), the loudness ratings of the various paths are evaluated: SLR , RLR and STMR for each end, and OLR in each direction.

The method is described in [24], but is not given in detail here. The loudness part of the model is important in its own right [for example in the study of Question 19/XII [25]], but not closely connected with the rest of the model. The program outputs loudness ratings calculated according to Recommendation P.79, but also calculates a set of loudness ratings according to the earlier method [26] which are used for subsequent calculations.

2.6 *Calculation of listening effort score*

This part of the model is intended to reproduce the result that would be obtained from a Listening Opinion Test.

It has been found possible to estimate $Y_{L\backslash dE}$ | by a process similar to those already well known in calculating loudness and articulation score. An intermediate quantity, Listening Opinion Index (LOI), is first calculated as follows. Each elementary band in the frequency range contributes to LOI an amount proportional to the product $B'_f P(Z_f)$, where B'_f is a frequency-weighting factor expressing the relative importance of that elementary band for effortless comprehension, and P is a growth function applied to the sensation level Z (which has already been evaluated for the loudness calculation). The actual values of the frequency-weightings differ somewhat from those used in loudness and articulation calculations; the growth function is limited to the range 0 to 1 as in articulation, but the form used is:

$$P(Z) = 10^{\frac{Z + 3.8}{10}} \quad \text{if } Z < -11,$$

$$P(Z) = 1 - 10^{\frac{-0.3(Z + 14)}{10}} \quad \text{otherwise.}$$

LOI is proportional to $\int B'_f \, P(Z_f) \, \mathrm{d}f$, but in practice the integral is replaced by a summation over a number of bands (normally 14), within each of which Z_f and B'_f are reasonably constant, just as in the loudness evaluation. The formula actually used is:

$$\mathrm{LOI} = AD \sum_i B'_i \, P(Z_i)$$

where

B'_i is the frequency weighting for the i th band, (shown diagrammatically in Figure 2-1),

Z_i is the mean Z in the i th band,

P is the appropriate growth function (illustrated in Figure 2-2),

A is a multiplier depending on the received speech level, with the value 1 for a small range of levels around the optimum but decreasing rapidly outside this range (see Figure 2-3 where the zero abscissa now corresponds to OLR = 8 dB (Recommendation P.XXE [20]) instead of 4 dB as previously),

D is a multiplier depending on the received noise level (ICN-RLR) with a value decreasing slowly from 1 at negligible noise levels towards 0 at very high levels (see Figure 2-4).

Figure 2-1, p.

Figure 2-2, p.

Figure 2-3, p.

Figure 2-4, p.

Thus it is only for wide-band, noise-free, distortion-free speech at optimum listening level that LOI attains its maximum value of unity.

The Listening Opinion Index is related to $Y_{L\backslash dE}$ | in a manner which depends on the standard of transmission to which listeners have been accustomed in their recent experience. It is found that the subjects' standard of judgement is influenced mostly by the best circuit condition experienced in the current experiment, or, in real calls, by the quality of the best connections normally experienced. For example, a circuit condition which earns a score of almost 4 in an experiment where it is the best condition, would earn a score of perhaps only 3 if a practically perfect condition were included in the same experiment, and about 3.5 if the best condition in the same experiment were equivalent in performance to the best connection that can normally occur in the British Telecom system. A parameter LOI LIM , introduced to cater for this effect, specifies the value of LOI that corresponds to maximum $Y_{L\backslash dE}$; it is generally set equal to 0.885 when connections are being judged against a background of experience with the British Telecom network. The relationship in general terms is

$$\ln\left(\frac{f I Y_{LE}\, f R}{- Y_{LE}\, f R}\right) = 1.465\left[\ln\left(\frac{OI}{OI_{LIM} - LOI}\right) - 0.75\right]$$

as shown in Figure 2-5. This brings us to the point where $Y_{L\backslash dE}$ | has been evaluated for each participant as a function of listening level — in particular, at the listening level established for each participant when the other speaks at Reference Vocal Level (RVL), defined in [27].

2.7 *Calculation of Conversation Opinion Score*

In order to convert a value of $Y_{L\backslash dE}$ | at the appropriate listening level to the corresponding value of Conversation Opinion Score (Y_C), it is necessary to take account of deviations of mean vocal level from RVL.

The symbol V_L | is used to denote the electrical speech level in dBV at the output of a sending end when the acoustic level at the input (mouth reference point) is RVL. During conversation, a different level (V_C) will generally prevail at the same point, because participants tend to raise their voices if incoming speech is faint or poor in quality and to lower them if incoming speech is loud. In other words, V_C at end *A* depends on $Y_{L\backslash dE}$ at end *A* , which depends on V_C at end B, which depends on $Y_{L\backslash dE}$ at end *B* , which depends in turn on V_C at end *A* . Thus there is a circular dependence or feedback effect.

Figure 2-5, p.

The sidetone paths introduce complications when STMR < | 3 dB (besides contributing noise from the environment to the receiving channel as already explained). Other things being equal, each talker's vocal level goes down by almost 1 dB for every 3 dB decrease in STMR below 13 dB, and this of course further modifies the opinion scores and speech levels at both ends by virtue of the feedback effect.

In addition to this, very high sidetone levels are experienced as unpleasant *per se* , particularly when the connection is poor for other reasons.

This complex interrelationship is found to be reasonably well represented by the following equations.

Y'_C is an intermediate quantity explained below.

$$\ln\left(\frac{fIY'_C\, fR}{-Y'_C\, fR}\right) = 0.7\left[\ln\left(\frac{fIY_{LE}\, fR}{-Y_{LE}\, fR}\right) + 0.5 - \frac{(13 - STMR)}{0}\left(\frac{-Y_{LE}\, fR}{fIY_{LE}\, fR}\right) \$\$2x\,2\right]$$

(2-1)

$$V_C - V_L = 4.0 - 2.1\,Y'_C - 0.3\,K\,(13 - \mathrm{STMR})$$

(2-2)

$$\ln\left(\frac{fIY_C\, fR}{-Y_C\, fR}\right) = 0.8451\,\ln\left(\frac{fIY'_C\, fR}{-Y'_C\, fR}\right) - 0.2727$$

(2-3)

where

K = 1 if STMR < 13,

K = 0 otherwise.

By substituting in equation (2-1) the value of $Y_{L\backslash dE}$ | already found for end A — which would apply for $V_C = V_L$ at end B — one obtains a first approximation to Y'_C, then from equation (2-2) an approximation to V_C at end A. The earlier calculations are repeated with this speech level to find a new value of $Y_{L\backslash dE}$ at end B, hence an approximation to Y'_C and V_C at end B. This process is repeated cyclically until each Y'_C converges to a settled value, and then equations (2-1) and (2-2) are simultaneously satisfied.

Figure 2-6 shows the form of the resultant relationship between $Y_{L\backslash dE}$ | and Y'_C, for two different values of STMR, with V_C at its proper value. The transformation [equation (2-3)], illustrated in Figure 2-7, is then applied to the intermediate score Y'_C, to give the estimated Conversation Opinion Score Y_c, which is shown as a function of $Y_{L\backslash dE}$ in Figure 2-8.

Figure 2-6, p.

Figures 2-7 and 2-8, p.

2.8 *Evaluation of other subjective measures of performance*

Relationships have been developed for various dichotomies of the opinion scale — such as proportion of votes greater than 2 (i.e. votes ‘‘Excellent’’ or ‘‘Good’’) — and for the percentage of positive replies to the ‘‘Difficulty’’ question (Supplement No. 2).

For example, percentage ‘‘Difficulty’’ is represented by the equation

$$\ln\left(\frac{f\,ID}{-D}\right) = -2.3\,\ln\left(\frac{fIY_C fR}{-Y_C fR}\right)$$

where

$D \times 100 = \%D$.

However, these relationships are satisfactory only for certain kinds of degradation and are still under review.

2.9 *Correspondence between calculated and observed values*

For symmetrical connections, provided very high sidetone levels and very high room noise levels are excluded, the model reproduces fairly well the results of laboratory conversation tests carried out in the U.K. In the most recent laboratory tests there is a tendency for speech levels and hence opinion scores to be somewhat lower than those observed earlier, but the relativities between circuit conditions are not much disturbed by this. It is believed, but not yet fully established, that approximately the same relativities hold good for other populations of subjects — in particular, for the population of ordinary telephone users accustomed to the British Telecom system — even though different absolute values of scores may be obtained from other populations of subjects or by using different experimental procedures.

Comparatively few results are available from experiments on asymmetrical connections, but such evidence as there is indicates that the model predicts too much divergence between the two ends of the connection — especially in respect of V_C, less so in respect of Y_C. It is proposed to introduce a feedback feature to reduce the divergence between the two V_C values, but care will be needed not to reduce the Y_C divergence too far as a result of this. HRC 4 in Annex A gives an example of CATNAP calculations for a set of connections with asymmetrical losses: compare these predictions with Reference [30] there quoted.

Predictions of Y_C | and V_C | from both CATNAP83 have been compared with the results of a number of conversation experiments conducted in the U.K. since 1976. The degree of agreement is summed up in Table 2-1.

H.T. [T2.3]

TABLE 2-1

Comparison of observed (O) and predicted (P) results for two models

Program	Types of connection	No. of conversations	Deviations (O — P)			
			Mean		r.m.s.	V
CATNAP	*Symmetrical only*	*680*	*—0.8*	*—0.29*	*4.1*	*0.41*
CATNAP	*Symmetrical and asymmetrical*	*883*	*—1.0*	*—0.22*	*3.8*	*0.38*
CATNAP83	*Symmetrical only*	*680*	*—0.2*	*—0.02*	*4.0*	*0.26*
CATNAP83	*Symmetrical and asymmetrical*	*883*	*—0.4*	*+0.14*	*3.8*	*0.44*

Table 2-1 [T2.3], p.

It will be seen that the improvement in Y_C | as predicted for symmetrical connections has been achieved at the cost of a slight increase in the r.m.s. deviation of Y_C when asymmetrical connections are included. But in view of the further alterations expected to be needed for the adequate prediction of the performance of asymmetrical connections, it is appropriate at the present stage to be guided mainly by the results for symmetrical connections.

2.10 *Incorporating miscellaneous degradations*

2.10.1 *PCM quantizing distortion*

Reference [28] describes a method for handling the effects of quantizing distortion in PCM systems. It is there established that a quantity Q, effective speech-to-quantization-noise ratio in dB, can be evaluated for any specified type of PCM system as a function of input speech level. It has been found that the subjective effect of a given value of Q can be approximated by that of a level of continuous circuit noise G dB below the speech level, where

$$G = 1.07 + 0.285\,Q + 0.0602\,Q^2.$$

Thus for a connection involving PCM links, one must include an evaluation of equivalent noise level in the iterative process that determines V_C: each successive approximation to V_C leads to a new value for Q, hence to a new value for G, and hence to a new contribution to the circuit noise to be taken into account in calculating the new value of $Y_{L\backslash dE}$. In practice these modifications have negligible effect unless the speech level at the input to the PCM system falls below about —25 dBV, or the circuit noise at the same point is very high, or the speech input level is so high (say > | (em5 dBV) that appreciable peak limiting occurs.

2.10.2 *Syllabic companding*

The case of a 2:1 syllabic compandor can be simply handled by finding a subjectively equivalent continuous noise level.

Let S | be the speech level at the input to the compressor, and N | be the noise level (psophometrically weighted) arising between the compressor and expander, both in dB relative to unaffected level. The resultant levels at the output of the expander will then be as given in Table 2-2.

H.T. [T3.3]
TABLE 2-2

	Speech	Noise while speech present	Noise while speech absent
Level at compressor input	S	—	—
Gain of compressor (dB)	—S/2	—	—
{ Level at compressor output and expander input }	S/2	N	N
Gain of expander (dB)	S/2	S/2	N
Level at expander output	S	N | | /2	2N
{ Level at same point in absence of compander }	S	N	N
Improvement	—	—S/2	—N

Table 2-2 [T3.3], p.

Note that S | and N | are both normally negative, so that the improvements are positive. Any noise present at the compressor input will be present at the same level at the expander output, and will combine by power addition with the other noise at the same point.

Subjectively equivalent performance is obtained by omitting the compandor and substituting a continuous noise level satisfying the condition:

Total improvement = 1/3 (improvement in presence of speech) +

+2/3 (improvement in absence of speech)

$= \quad —S/6 — 2N/3.$

Hence

equivalent noise level = N — improvement

$= \quad N + S/6 + 2N/3 = S/6 + 5N/3.$

This noise level is recalculated from V_C | on each iteration and used to calculate the next value of $Y_{L\backslash dE}$.

2.10.3 *Delay and echo*

The audibility and objectionability of echo can be expressed as a reasonably simple function of the delay and loudness rating of the echo path, but the wider effects of echo and main-path delay in disrupting conversation can at present only be treated by *ad hoc* estimation from the known performance of circuit conditions in neighbouring parts of the range. Steps are being taken to extend the model in this direction, account being taken also of the interaction of delay and echo with sidetone and nonlinear distortions.

2.10.4 *Crosstalk*

The loudness part of the model may be used to estimate the audibility of crosstalk, at various attenuations, and hence to find the attenuation required to reduce it to an inaudible level or to an acceptable level.

2.11 *Practical use of the model*

At the academic or research level, the chief use of a model of this kind is in promoting an understanding of the fundamentals of telecommunication between human beings, and in finding potential improvements in the techniques of telecommunication systems.

At the practical level, the chief advantage of having the model available is that it encodes the knowledge of the performance of telephone connections in a very economical manner, obviating the need for large and complex tabulations or graphs. For connections containing only the ‘‘natural’’ degradations, the program CATNAP greatly facilitates routine use of the model. The user of this program need not know anything about the theory beyond the meaning of the terms and symbols used, and need not normally make any special measurements. Connections are specified in terms of standard items and quantities, such as noise levels, telephones of particular types, lengths of cable with stated resistance and capacitance per kilometre, and attenuators with stated loss. Starting from these data, the program performs all the necessary calculations and prints out loudness ratings, speech levels, and opinion scores ($Y_{L\backslash dE}$ and Y_C). More detail can be printed on request.

It would of course be possible to construct a large table of results covering a wide range of connections, but the table would have to be either too large to be practical or else limited by making arbitrary fixed choices for many of the variables. In either case the advantage of having the model — that it holds the information in an economically coded form and releases only the required part on demand — would be lost.

CATNAP may also be used inversely. Suppose it is desired to find what value of some variable in a connection (the independent variable) will yield a given value of one of the dependent variables. By performing runs at different values of the independent variable one identifies a region within which the required value lies; one can then repeat the calculation at ever smaller intervals until the required value is located with sufficient accuracy. For example, where all features except the local line remain fixed, one can find the line length (for the type of cable in question) that will yield values of OLR below some specified maximum, or values of Y_C above some specified minimum. More than one independent variable could of course be adjusted, but correspondingly more work would then be needed in order to find the combinations that satisfied the criterion.

The usefulness of these facilities is evident.

3 Calculation of transmission performance from objective measurements by the information index method (Contribution by France)

3.1 *Introduction; type of model*

The information index theory is given in [31]. This quantity can be calculated from the results of objective measurements and some fundamental data on speech and hearing. Among the factors listed in Question 7/XII [32] the theory takes into account transmission loss, circuit noise, room noise, attenuation/frequency distortion, sidetone and various distortions occuring in digital transmission (Question 18/XII). The effect of other types of nonlinear distortion is under study.

The model used here belongs to the second type mentioned in [33] and in § 2.1 of this Supplement, since it reflects de cause-and-effect relationships between the input (properties of the connection considered, acoustic environment, loudness properties of speech and hearing) and the output (mutual information transmitted between speaker and listener). This Section only describes the practical method for performing the computation of the information index. As shown in [31] and also in Tables 3-4 and 3-7 below, the values thus computed are strongly correlated with the results of subjective opinion tests carried out in several countries.

3.2 *Application to digital transmission*

3.2.1 *Definitions*

Table 3-1 defines the various signal-to-noise ratios to be considered (in dB).

H.T. [T4.3]

TABLE 3-1

Notations See Note 1	See Note 2	Definitions
Q *Signal-to-noise ratio, kept constant by a MNRU* *}*	*Q* , *Q*	*{*
Q *Segmental signal-to-noise or signal-to-distortion ratio (in dB) (mean of ratios computed over segments of 16 or 32 ms)* *}*	*Q*	*{*
Q *Ratio (in dB) of the mean signal power to mean noise or distortion power, for speech-correlated noise* *}*		*{*

Note 1 — Over the transmitted band.

Note 2 — At frequency *f* .

Table 3-1 [T4.3], p.

Let *s* | be the original speech signal and *r* | the reconstructed signal, we have:

$$Q_p = 10 \log_{10} \left[\int s^2 / \int (s - r)^2 \right] \text{ dB} \qquad (3\text{-}1)$$

If the sums are taken over an entire speech utterance, Q_p | is not a satisfactory quality criterion; for a sampling frequency of 8 kHz, we have:

Q_{seg}

[Unable to Convert Formula]

$$\sum_{m=0}^{M-1}$$

$$10 \log_{10} \left\{ \frac{\sum_{j=1}^{128} s^2(j + 128m)}{\sum_{j=1}^{128} [s(j + 128m) - r(j + 128m)]^2} \right\} \text{ dB} \tag{3-2}$$

where M | is the number of 16 ms segments.

To determine Q_s, the spectra of the signal, s | and of the distortion (s — r) are computed over 256 samples of 32 ms duration and divided into the appropriate frequency bands. Then the segmental signal-to-distortion ratio is computed in each band.

3.2.2 *Basic formulas*

The information index I_I | (in dB), defined in [31], is given by

$$I_I = \sum_j B_j \times V_j \tag{3-3}$$

with

V_j =
[Unable to Convert Formula]

(3-4)

B_j | is the weight allocated to the band of rank j ; C_j = 10 log$_{1\backslash d0}$($f_j/\Delta f_c$), Δf_c | being the critical bandwidth

Table 3-2 gives the values of B_j | and C_j | for the bands which are used in the example of § 3.2.4; they are reproduced in lines 70 and 80 of Appendix I. Values for ISO preferred frequencies (3rd octave spaced) from 0.1 to 8 kHz are given in lines 180-370 of Appendix II under columns BJ and CJ.

H.T. [T5.3]
TABLE 3-2
Frequency weighting

j	Equal articulation bands { Extreme frequencies (Hz) }		$B \times 10^5$	C (dB)
1	*| 00*	*| 30*	*5 | 57*	*4.1*
2	*| 30*	*| 30*	*4 | 33*	*5.6*
3	*| 30*	*| 60*	*6 | 82*	*6.4*
4	*| 60*	*| 00*	*7 | 97*	*6.9*
5	*| 00*	*| 40*	*6 | 46*	*7.4*
6	*| 40*	*1 | 00*	*6 | 22*	*7.8*
7	*1 | 00*	*1 | 50*	*5 | 85*	*8.0*
8	*1 | 50*	*1 | 10*	*5 | 00*	*8.0*
9	*1 | 10*	*1 | 80*	*5 | 73*	*8.2*
10	*1 | 80*	*1 | 60*	*5 | 17*	*8.2*
11	*1 | 60*	*1 | 30*	*4 | 17*	*8.2*
12	*1 | 30*	*2 | 20*	*4 | 06*	*8.2*
13	*2 | 20*	*2 | 40*	*5 | 73*	*8.2*
14	*2 | 40*	*2 | 00*	*5 | 61*	*8.2*
15	*2 | 00*	*2 | 20*	*6 | 10*	*8.2*
16	*2 | 20*	*3 | 00*	*6 | 86*	*8.1*
	TOTAL	*102 | 58*		

Table 3-2 [T5.3], p.

3.2.3 *Relations between signal-to-noise ratios in the case of digital transmission*

In the case of MNRUs with uniform or shaped noise, from the very principle of their operation, $Q_s = Q_j$ and Equation (3-4) may be applied directly if Q_s in each band is known.

For digital coders, the equivalence law in lines 150-170 of Appendix I is used. The law depends on two parameters, K_1 and $d = Q_{s\backslash de\backslash dg} — Q_p$. Numerical computations have shown that this law is valid both for PCM ($K_1 = 5.2$; $d = 0$) and for natural speech ($K_1 = 5.2$; $d = —5.3$) [31]. The example of § 3.2.4 shows that it gives consistent results for various types of coders.

3.2.4 *Program and example of application*

The program used is reproduced in Appendix I.

Table 3-3 gives measured values of the signal-to-noise ratios defined above for MNRUs and for a variety of codecs, as well as the information index values computed from these results as the mean opinion scores (MOS) for listening determined in the CNET Laboratory [34].

Table 3-4 shows the correlation of these MOS with the information index (Table 3-3) and with other objective measures of transmission performance which have been proposed.

3.3 *Application to analogue transmission*

3.3.1 *General; use of the program*

The calculation of the information index, in the case of analogue transmission, will be explained with reference to the program reproduced as Appendix II. This applies to a connection composed of two telephone sets of the NTT 600 type (with 7 dB subscriber lines), one SRAEN filter and a variable attenuation. Writing the corresponding program for other types of connection is discussed in § 3.4.

The program is used in the following way:

a) enter RN, STMR, ICN_0 | as defined in lines 30-60, press ‘‘L’’, enter OLR; read IN = information index (listening);

b) if I_c | (information index under conversation conditions) is required, press ‘‘C’’; read IN = I_c;

c) press ‘‘T’’.

3.3.2 *Data*

Lines 170-370 of the program.

Lines 180-370 correspond to 1/3 octave spaced frequencies from 0.1 to 8 kHz.

3.3.2.1 *Basic data*

These do not depend on the type of telephone set used.

BK = Hearing threshold for continuous-spectrum sounds (L_s in [31]) referred to ear reference point;

S = Spectrum density (long-term mean intensity) of speech at the mouth reference point; S + 0.4 dB corresponds to a vocal level —4.7 dB/1 Pa;

See Table 3-5.

BJ = frequency weighting (see [31]);

CJ = correction term in formula 3.4 giving V_j.

3.3.2.2 *Electroacoustic characteristics*

These depend on the connection considered.

SRL = Loss (send + receive) of the local system;

D1 = Loss of the line filter.

H.T. [T6.3]

[Unable to Convert Table]

MNRUS = MNRU with shaped noise

F = ADPCM with fixed predictor

V = ADPCM with variable predictor

SB = Sub-band coding

Table 3-3 [T6.3], A L'ITALIENNE, p.

H.T. [T7.3]

TABLE 3-4

{

Correlation between MOS and various objective measures of transmission performance in the case of digital transmission

}

Objective measure (Note)	Systems									
	Group B PCM, ADM, ADPCM-F { { {		{							
	R	S	R	S	R	S	R	S	R	S
Q (SNR)	*0.798*	*0.578*	*0.803*	*0.559*	*0.687*	*0.680*	*0.590*	*0.711*	*0.650*	
Q (SNR seg)	*0.950*	*0.301*	*0.894*	*0.430*	*0.906*	*0.396*	*0.725*	*0.606*	*0.720*	
''Log likelihood ratio''	*0.943*	*0.213*	*0.924*	*0.341*						
Cepstrale distance	*0.954*	*0.208*	*0.929*	*0.331*						
SRNF									*0.884*	
Information index					*0.994*	*0.101*	*0.993*	*0.102*	*0.976*	*0.175*

Note — Notations from Table 3-1; see also Contribution COM XII-No. 8 (Study Period 1985-1988).

SNRF = *Q*
| *frequency weighted*

R = Correlation coefficient

S = Standard deviation (in terms of MOS on a 0 to 4 or 1 to 5 scale)

Table 3-4 [T7.3] A L'ITALIENNE, p.

3.3.2.3 *Noise components*

The following components (which depend on the connection) are considered.

BDFE = spectrum of far-end room noise via far-end telephone set;

BDCN = spectrum of circuit noise;

BDST = spectrum of near-end room noise via sidetone path;

BDEL = spectrum of near-end room noise via earcap leakage.

The data at lines 180-370 correspond to a typical connection. They are:

FE = BDFE computed for RN = 50 dBA and an overall loudness rating (OLR), according to Recommendation P.79, of 5 dB;

CN = BDCN computed for ICN_0 = —60 dBmp;

ST = BDST computes for RN = 50 dBA and STMR = 15 dB;

EL = BDEL computed for RN = 50 dBA.

The computations were made [35] from the frequency characteristics given in [36], by a method similar to that used for deriving Table 3 from Table 2 in [33].

3.3.3 *Computation of signal-to-noise ratios*

3.3.3.1 *Level of the signal*

First, OLR is corrected if it is smaller than the optimum value (see Appendix II, lines 100-160). This optimum is determined by a subroutine (lines 720-810) which is similar to the formulas of [38], but was adapted to the results of subjective opinion tests published in [36].

3.3.3.2 *Signal-to-noise ratio* | (lines 425-440)

The power sum of the noise components is taken and the signal-to-noise ratio Z_n thus obtained.

3.3.3.3 *Effect of thresholds* | (lines 450-480)

Z_a | is computed (see [31]) from which the equivalent signal-to-noise ratio Z_e is derived. The resultant Z is obtained by power summation of the noises corresponding to Z_n and Z_e.

3.3.4 *Information index for a constant speech level,* I_L

The equivalence between Z | and Q | is derived from the values under ‘‘Japan’’ in Table 1 of [37], then V | is computed at each frequency (lines 650-700) and IN for listening is obtained (lines 500-550).

3.3.5 *Conversation information index ,* I_c

First, speech power is modified to take into account the effect of sidetone when talking (lines 90 and 560-610), as in § 2.7 above.

A second correction is added (line 620), as explained in [31]. The application of the present model to 13-2P-27-type telephone sets with the equivalence law mentioned under 3.3.4 gives:

$$V_c - V_L = 9.87 - 0.4085\, I_L$$

3.3.6 *Examples*

Table 3-5 gives the MOS determined subjectively in two tests (one listening, one under conversation conditions) for the same conditions, reported in [36], and the information indexes computed for these conditions.

Table 3-6 gives the subjective MOS determined for various conditions of noise and the corresponding listening information indexes.

Table 3-7 shows the correlation between subjective MOS and the values of information index given in Table 3-5 and Table 3-6, as well as the results of similar calculations for 13-2P-27-type telephone sets.

H.T. [T8.3]
TABLE 3-5
**Information index I for NTT 600-type telephone sets (7 dB line) with SRAEN filter,
STMR = 7.1 dB and opinion scores from tests 2 and 6**

RN (dBA)	CN (dBmp)	ICN 0 (dBmp)	OLR (dB)	*Y*	*I* (dB)	*Y*	*I* (dB)
(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
60	*—62.1*	*—58.2*	*1.4*	*3.13*	*23.48*	*2.94*	*23.27*
			11.4	*2.5*	*22.75*	*2.34*	*22.50*
			21.4	*2.31*	*19.44*	*1.58*	*19.52*
			31.4	*0.65*	*12.28*	*0.2*	*15.13*
60	*—59.8*	*—55.9*	*1.4*	*3.1*	*23.52*		
			11.4	*2.91*	*22.73*		
			21.4	*1.75*	*19.35*		
			31.4	*0.8*	*12.02*		
60	*—55.8*	*—51.9*	*1.4*	*2.83*	*23.59*	*2.99*	*23.38*
			11.4	*2.75*	*22.65*	*2.39*	*22.39*
			21.4	*1.79*	*19.03*	*1.28*	*19.20*
			31.4	*0.5*	*11.24*	*0.43*	*14.71*
60	*—51.4*	*—47.5*	*1.4*	*3.06*	*23.66*	*3.08*	*23.43*
			11.4	*2.24*	*22.44*	*2.17*	*22.16*
			21.4	*1.05*	*18.27*	*1.29*	*18.64*
			31.4	*0.09*	*9.3*	*0.22*	*14.11*
60	*—45.6*	*—41.7*	*1.4*	*2.31*	*23.64*	*2.63*	*23.37*
			11.4	*1.4*	*21.67*	*1.73*	*21.39*
			21.4	*0.64*	*16.2*	*0.77*	*17.23*
			31.4	*0.05*	*5.57*	*0.13*	*12.08*

Explanation of columns

(1) Room noise, dBA

(2) Circuit noise at input to receiving end, dBmp

(3) ICN 0 = CN + 3.9 dB

(4) OLR (Rec. P.79)

(5) Listening MOS (on a 0 to 4 scale), test 2 of [36], p. 4-4

Listening information index (position L of Appendix II)

Conversation MOS, test 6 of [36], p. 4-9

Conversation opinion index (position C of Appendix II)

Tableau 3-5 [T8.3], p. 32

Blanc

H.T. [T9.3]
TABLE 3-6
Information index at listening for NTT 600-type telephone sets (7 dB line) with SRAEN filter, STMR = 7.1 dB and listening opinion score from test 4

RN (dBA)	CN (dBmp)	ICN 0 (dBmp)	OLR (dB)	*Y*	*I (dB)*
0	. (see Note)	{			
—100.					
(see Note)					
}	{				
—3.6					
1.4					
6.4					
11.4					
16.4					
21.4					
26.4					
31.4					
}	{				
2.30					
2.83					
3.26					
2.92					
2.59					
2.12					
1.89					
1.23					
}	{				
22.87					
23.40					
23.85					
23.55					
23.05					
22.45					
21.73					
20.91					
}					
60	—55.8	—51.9	{		
—3.6					
1.4					
6.4					
11.4					
16.4					
21.4					
26.4					
31.4					
}	{				
2.61					
2.94					
3.00					
2.38					
1.80					
1.41					
0.91					
0.44					
}	{				
22.78					
23.59					
23.51					
22.65					
21.24					
19.03					
15.95					
11.44					
}					
60	—56.9	—53.9	1.4 11.4 21.4 31.4	3.20 2.53 1.24 0.24	23.57 22.68 19.15 11.51
50	—55.8	—51.9	1.4 11.4 21.4 31.4	3.21 2.64 1.58 0.35	23.88 23.24 21.04 15.94

45 23.77 23.24 }	—64.9	—61.**9**	1.4 13.4	3.23 2.62	{

Note — In these cases, there was no circuit noise during the opinion tests but a noise corresponding to CN = —76.9 (ICN 0 = —73) is used in the OPINE model. An arbitrary low noise value may be used for the calculation of the information index.

Tableau 3-6 [T9.3], p. 33

Blanc

H.T. [T10.3]

TABLE 3-7
{
Correlation between MOS and the information index in the case of analogue transmission
}

Type of connection	{ Type of MOS RN (dBA)	Correlation coefficient ICN 0 (dBmp)	{ OLR (dB)				
{	0 to 60 —3.6 to +31.4 60	{ Y L —58.2 to —41.7	0.978 +1.4 to +31.4	0.15 Y C	—0.34 + 0.31 0.977	0.16	—0.32 + 0.32
{	50 { Y C	{ 0.995	0.07	—0.17 + 0.16			

Note — Notations are of Appendix II.

Tableau 3-7 [T10.3], A L'ITALIENNE, p. 34

3.4 *Possible extensions*

3.4.1 *Frequency characteristics*

Appendix II gives an example which is explained in § 3.3 above. If different types of sets, balancing networks, subscriber's lines or line filters are used, the corresponding data in Appendix II should be changed accordingly and the noise data recalculated. This is the same procedure as is given in the model of § 2.4 above and is explained in [33].

OLR and STMR, used as independent variables, should be recalculated according to Recommendation P.79.

3.4.2 *Connection including digital processes*

Paragraph 3.2 above and Appendix I apply to cases where speech is near its optimum level, in order to compare different coders under such conditions. If the coders give rise to appreciable clipping, the loss of information due to this effect should be calculated and the corresponding value of Q determined as explained in [31].

Anyhow, when digital process are included in a connection of a telephone network, the corresponding values of Q_m should be determined in each frequency band and combined with the value of Q in Appendix II, by a power summation of the noises and distortions.

MONTAGE : §.4 SUR LE RESTE DE CETTE PAGE