INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Supplement 3
(Series P)
(03/93)

TELEPHONE TRANSMISSION QUALITY

MODELS FOR PREDICTING TRANSMISSION QUALITY FROM OBJECTIVE MEASUREMENTS

Supplement 3 to
ITU-T Series P Recommendations

(Previously "CCITT Recommendations")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

Supplement 3 to ITU-T Series P Recommendations was revised by the ITU-T Study Group XII (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms “CCITT, CCIR or IFRB” or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

2 In this Recommendation, the expression “Administration” is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

CONTENTS

INTRODUCTION

Models for predicting the subjective opinion of telephone connections, using data from objective measurements, are currently under study. 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 clauses 1, 2, 3 and 4 below have been proposed by several Administrations.

Supplement 3

MODELS FOR PREDICTING TRANSMISSION QUALITY FROM OBJECTIVE MEASUREMENTS

(referred to in Series P Recommendations)

1 Transmission rating models *(Geneva, 1980; modified at Malaga-Torremolinos, 1984) (Quoted in clause 3/P.11)* (Contribution by the United States of America and Canada)

1.1 Introduction

This clause 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 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 five-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 [3]. Subsequent test results led to an alternative form of the model. 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 [6], [7], [8].

The model for bandwidth and attenuation distortion is based on tests conducted in 1978 [9].

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 [10].

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 [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. An important reason for the introduction of this scale was the 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.

TABLE 1-1

Overall loudness rating (dB)	Circuit noise (dBmp)[a]	Transmission rating
16 31	−61 −76	80 40
[a] The circuit noise values are referred to a receiving system with a receiving loudness rating (RLR) = 0 dB.		

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 clause 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).

NOTE – In equation (1-1) the value of L'_e can be replaced by $L'_e + L_s$ to provide a loudness loss correction in (L_s) in dB. This compensates for reduced talker speech level when the sidetone masking rating (STMR) at the talker end of the connection is less than 15 dB. The correction L_s is zero when STMR = 12, the default value. Otherwise, the correction L_s is given as follows:

$$L_s = -0.3\ (\text{STMR} - 12) \quad \text{if STMR} < 15$$
$$L_s = -0.9\ (\text{STMR} - 12) \quad \text{if STMR} \geq 15$$

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 $N'_{Re} = -64$ dBmp and the adjustment to R_{LN} as recommended in the Note in 1.2.2, rather than the value of $N'_{Re} = -58.63$ used in Volume V of the *Blue Book*.

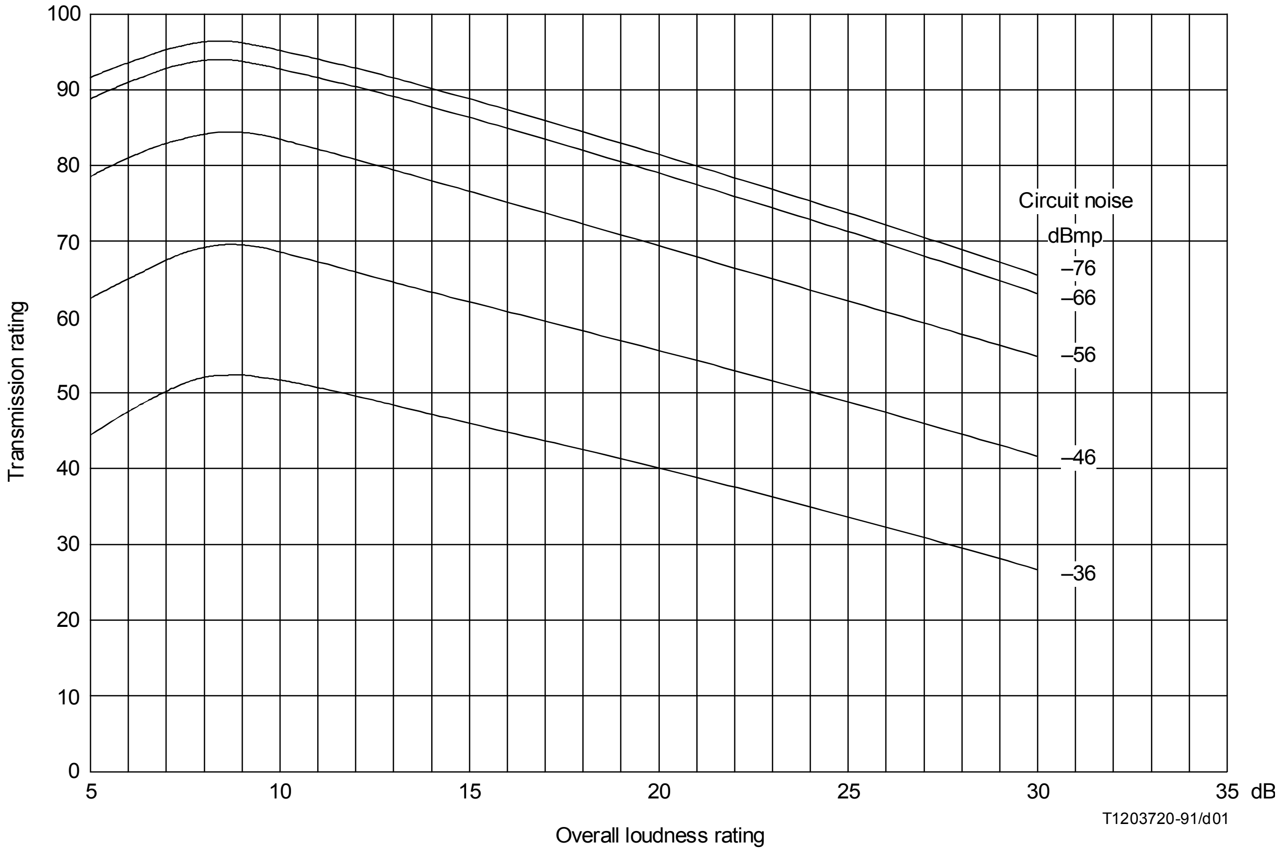


FIGURE 1-1

Transmission rating for OLR and noise

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

$$N'_{Re} = N_R - 121 + 0.0078\,(N_R - 35)^2 + 10 \log_{10}\left[1 + 10^{\frac{6 - L'_s}{10}}\right] \qquad (1\text{-}2)$$

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'_{Re}, 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

$$N'_{Re} = -58.63 \text{ dBmp} \tag{1-3}$$

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.

However, North American tests conducted in 1987 indicate that a lower value for N'_{Re} is more appropriate. The value which has been agreed in North America is

$$N'_{Re} = -64 \text{ dBmp}$$

When this lower value is used, the value of R_{LN} should be replaced with R'_{LN} as follows:

$$R'_{LN} = -0.0023\,(R_{LN})^2 + 1.21\,R_{LN} - 4.7$$

This adjustment provides realistic opinion results when used with Murray Hill database. Although values of N'_{Re} less than -64 dBmp may be useful in fitting test results from laboratory tests with low room noise, they are not recommended for use in predicting subjective opinion results.

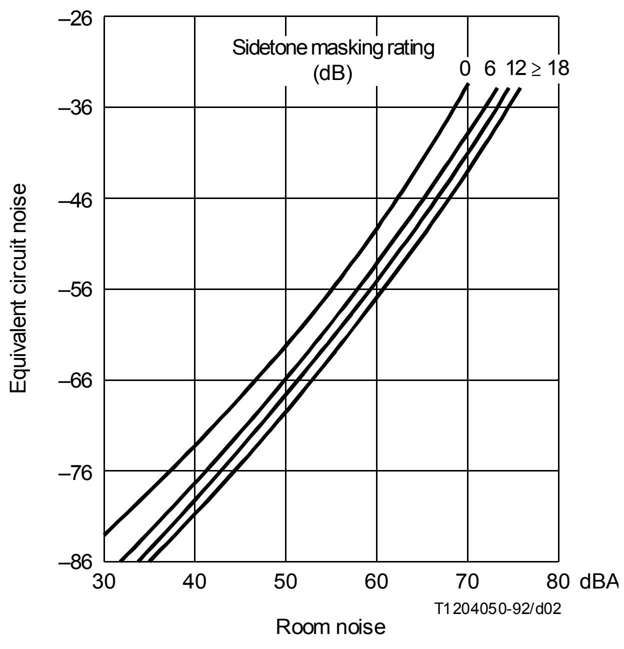


FIGURE 1-2

Equivalent circuit noise for room noise

1.2.3 Circuit noise equivalent of quantizing noise

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

$$N'_{Qe} = V - 2 - SNR \tag{1-4}$$

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

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

from which

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

Based on a 1975-1976 Speech Level Survey, [11] the speech level for domestic North American connections can be approximated by

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

from which

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

Estimates of Q for single codec pairs are given below for Pulse Code Modulation (PCM), Nearly-Instantaneous Compandored modulation (NIC), Adaptive Differential Pulse Code Modulation (ADPCM) and Adaptive Delta Modulation (ADM). They apply to the particular algorithms described in [6] and [8].

$$\text{PCM:} \quad Q = 0.78\,L - 12.9 \tag{1-8}$$

$$\text{NIC:} \quad Q = 0.74\,L - 2{,}8 \tag{1-9}$$

$$\text{ADM:} \quad Q = 0.42\,L + 8.6 \tag{1-10}$$

$$\text{ADPCM:} \quad Q = 0.98\,L - 5.3 \tag{1-11}$$

$$\text{ADPCM-V:} \quad Q = 1.04\,L - 4.6 \tag{1-12}$$

where

L is the line bit rate in kbit/s.

NOTE – The ADPCM algorithm with fixed predictor is described in [12]. The ADPCM-V algorithm with adaptive predictor is described in [8].

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

$$Q = -15 \log_{10}\left[\sum_{i=1}^{n} 10^{-\frac{Q_i}{15}}\right] \tag{1-13}$$

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}, for overall loudness rating, circuit noise and bandwidth is

$$R_{LNBW} = (R_{LN} - 22.8)\, k_{BW} + 22.8 \tag{1-14}$$

where

$$k_{BW} = k_1\, k_2\, k_3\, k_4 \tag{1-15}$$

with

$$k_1 = 1 - 0.00148\,(F_l - 310) \tag{1-16}$$

$$k_2 = 1 + 0.000429\,(F_u - 3200) \tag{1-17}$$

$$k_3 = 1 + 0.0372\,(S_l - 2) + 0.00215\,(S_l - 2)^2 \tag{1-18}$$

$$k_4 = 1 + 0.0119\,(S_u - 3) - 0.000532\,(S_u - 3)^2 - 0.00336\,(S_u - 3)\,(S_l - 2) \tag{1-19}$$

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_{BW}. Figure 1-4 makes use of the expression for k_3k_4 in Note 2 below.

NOTES

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

2 In accordance with test results conducted in 1987 and with little change in the predicted results the product k_3k_4 can be replaced with

$$k_3k_4 = 0.93 + 0.0627\,(S_l + 0.441\,S_u) - 0.00012\,(S_l + 0.441\,S_u + 7.17)^3$$

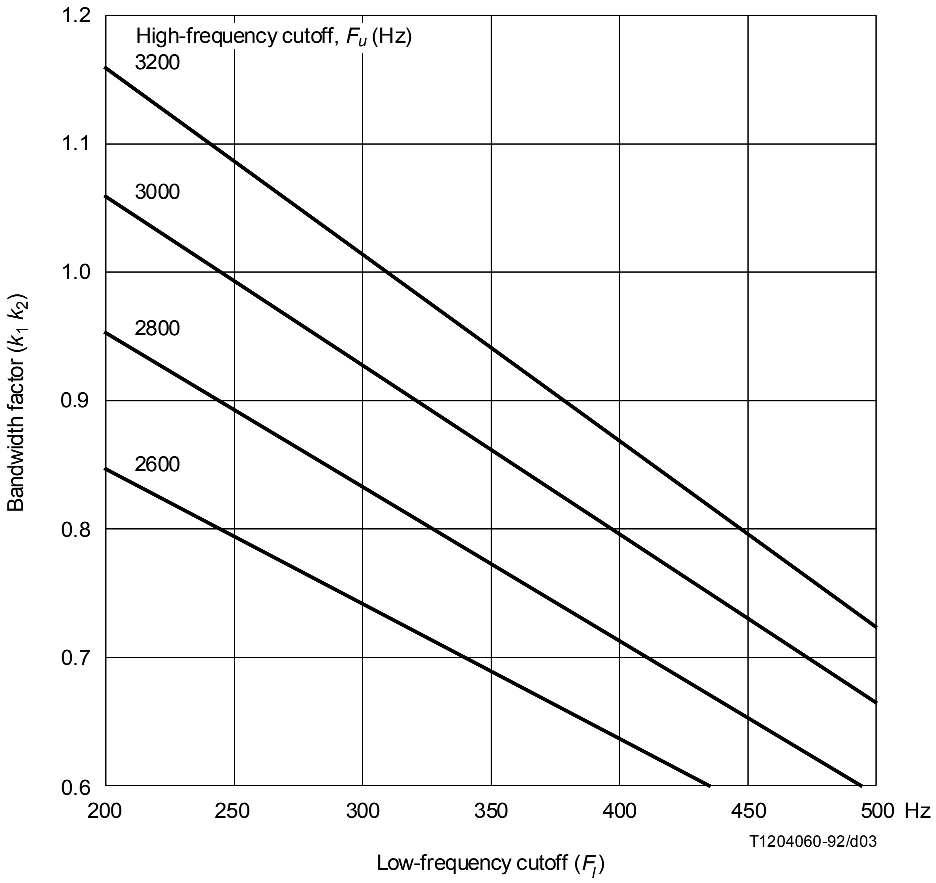


FIGURE 1-3

Bandwidth model factor

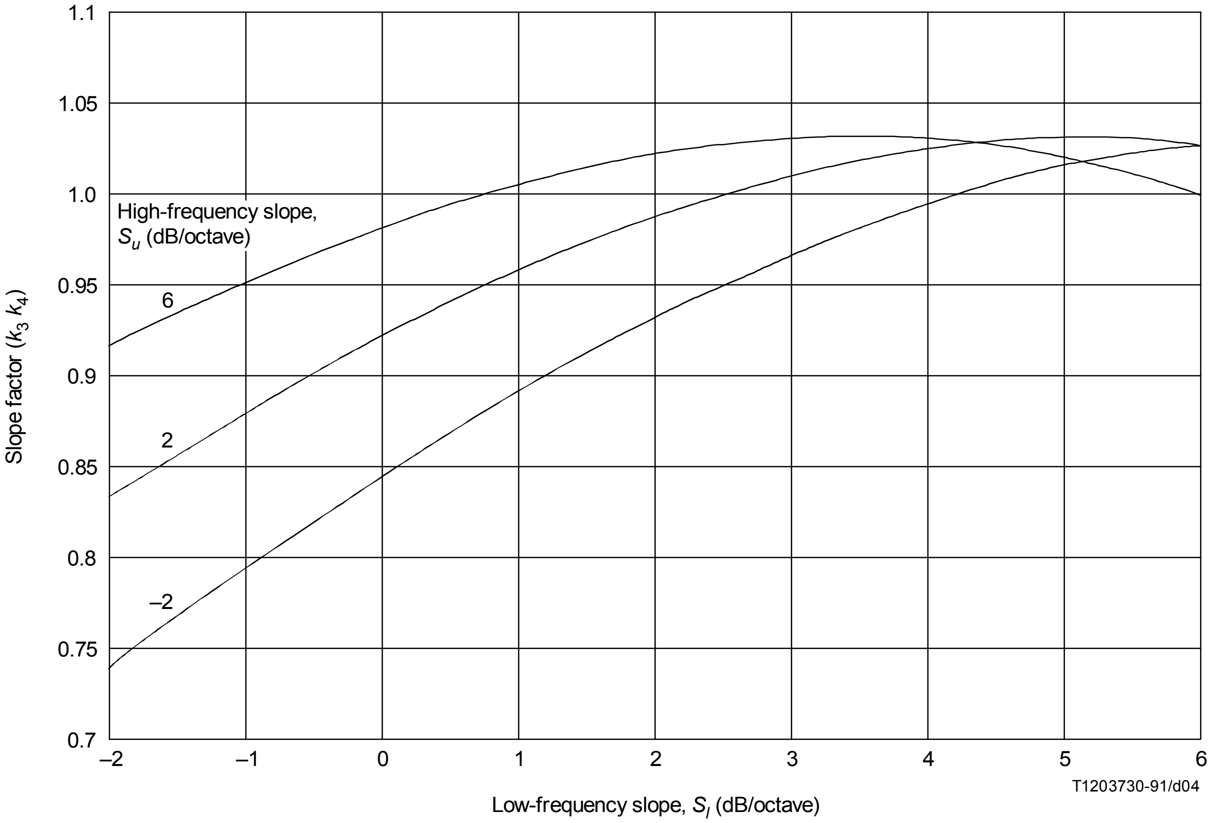


FIGURE 1-4

Attenuation – Distortion model factor

1.2.5 Listener echo

The transmission rating model for listener echo is

$$R_{LE} = 9.3\,(WEPL + 7)\,(D_L - 0.4)^{-0.229} \qquad (1\text{-}20)$$

where

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

$$WEPL = -20\log_{10}\frac{1}{3200}\int_{200}^{3400} 10^{-\frac{EPL\,(f)}{20}}\,df \qquad (1\text{-}21)$$

$EPL(f)$ 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_{LE}, as a function of the weighted echo path loss and listener echo-path delay is shown in Figure 1-5.

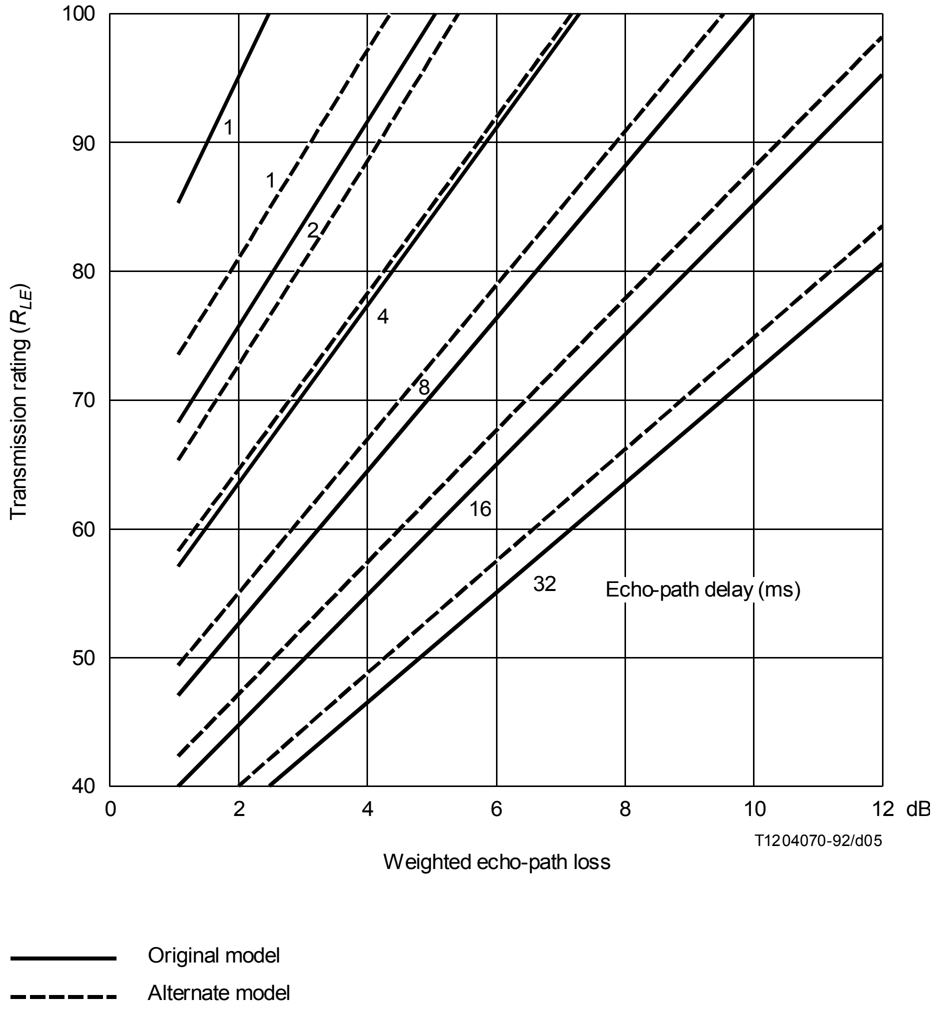


FIGURE 1-5

Transmission rating for listener echo

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:

$$R_{LNLE} = \frac{R_{LN} + R_{LE}}{2} - \sqrt{\left[\frac{R_{LN} - R_{LE}}{2}\right]^2 + 13^2} \qquad (1\text{-}22)$$

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.

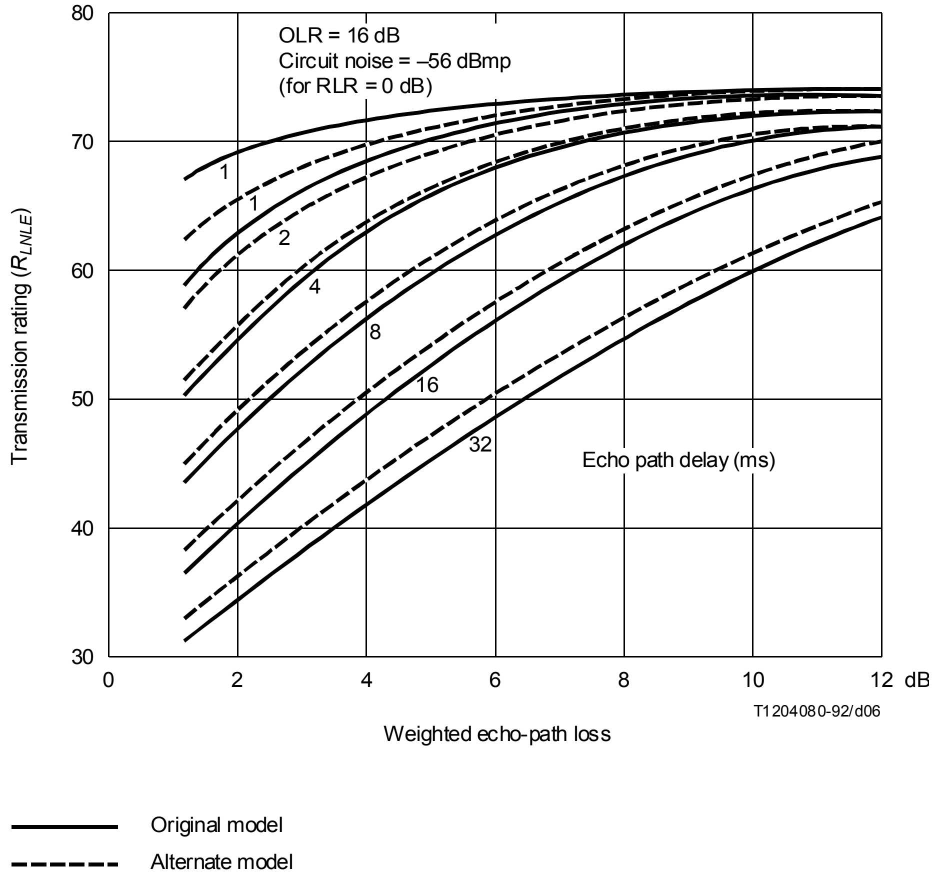


FIGURE 1-6

Transmission rating for OLR, circuit noise and listener echo

NOTE – The preceding material is based on the use of a specific set of test results and the listener echo model of [3]. Subsequently, new test results were reported in [4] and [5] 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 [5] 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

$$SWEL = SM + SF\frac{SM}{1 + SM}$$

Hence, like *WEPL*,

$$SWEPL = SM, \text{ if } SF = 0$$

Also,

$$SWEPL \approx WEPL, \text{ for } SM >> 1$$

The effect of the shape factor is reduced as *SM* approaches 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

$$R_E = 92.73 - 53.45 \log_{10}\left[\frac{1 + D}{\sqrt{1 + \left(\frac{D}{480}\right)^2}}\right] + 2.277\, E \qquad (1\text{-}23)$$

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_{LN} model.

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

$$R_{LNE} = \frac{R_{LN} + R_E}{2} - \sqrt{\left(\frac{R_{LN} - R_E}{2}\right)^2 + 100} \qquad (1\text{-}24)$$

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_{LN-ST}, for OLR, total effective noise and sidetone is

$$R_{LN\text{-}ST} = K_{ST}\, R_{LN} \qquad (1\text{-}25)$$

and for talker echo and sidetone is

$$R_{E-ST} = R_E + [2.6\,(12 - SL) - 1.5\,(4.5 - SR)^2 + 3.38]\,[D_0^2 \,/\, (D_0^2 + D^2)] \qquad (1\text{-}26)$$

where $D_0 = 30$ and D is defined with equation (1-23).

The sidetone factor, K_{ST}, is calculated from

$$K_{ST} = 1.00 + C\,[0.021 - 0.002\,(SL - 15)^2 + 0.001\,(SR - 2)^2\,(SL - 15)] \qquad (1\text{-}27)$$

where C is a coefficient which determines the predicted magnitude of the sidetone effect. A value of C = 1 corresponds to the original test data. However, a smaller value of C (for example C = 0.25) seems to provide better agreement with some test results from other countries. Provisionally a value of C = 0.25 is suggested for use.

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

Figure 1-9 shows curves obtained by determining R_{LN-ST} and R_{E-ST}, then substituting these values for R_{LN} and R_E respectively in equation (1-24). This figure uses a value of C = 1 in equation (1-27) for K_{ST}. A lower value of C will produce higher values of transmission rating at the lower values of sidetone masking rating.

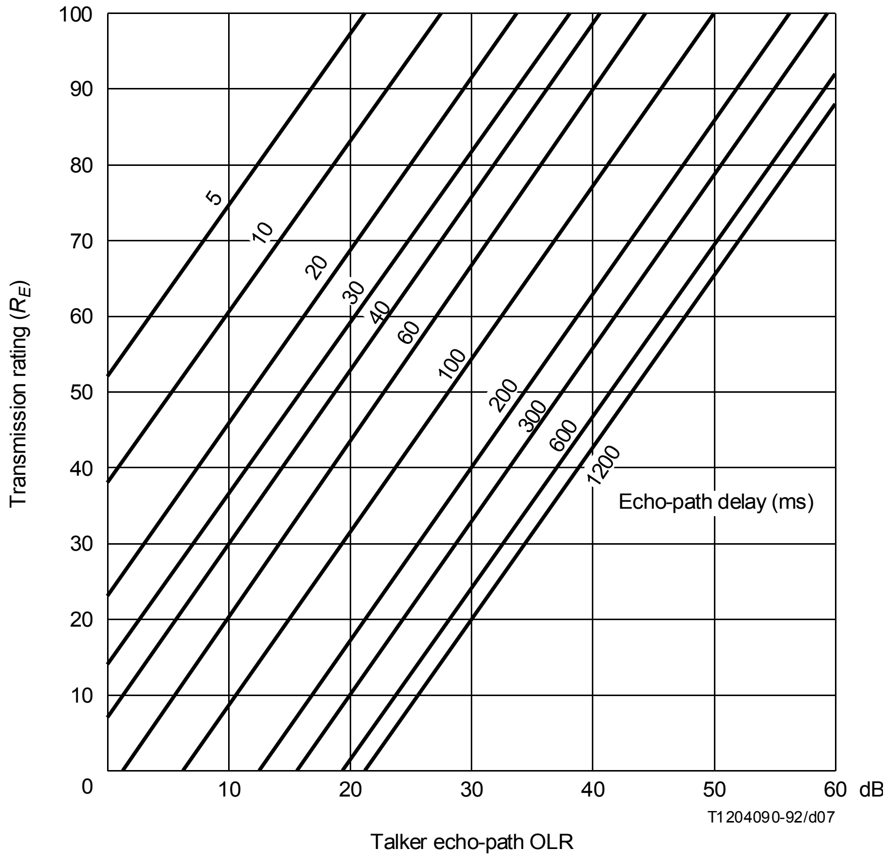


FIGURE 1-7

Transmission rating for talker echo

---

[1] *Sidetone Response:*

Below 1 kHz	*Above 1 kHz*
0	0
+3.0	+4.5
+6.0	+9.0

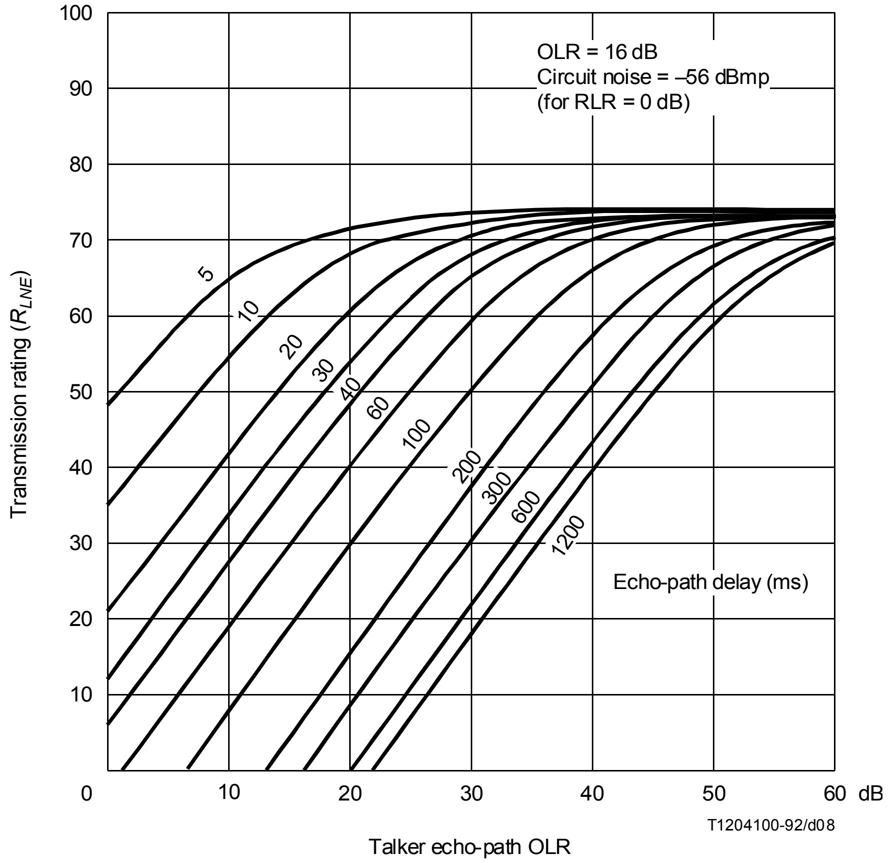


FIGURE 1-8

Transmission rating for OLR, circuit noise and talker echo

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:

$$G + E = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{A} e^{-\frac{t^2}{2}} \, dt \tag{1-28}$$

$$P + U = \frac{1}{2\pi} \int_{B}^{-\infty} e^{-\frac{t^2}{2}} \, dt \tag{1-29}$$

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

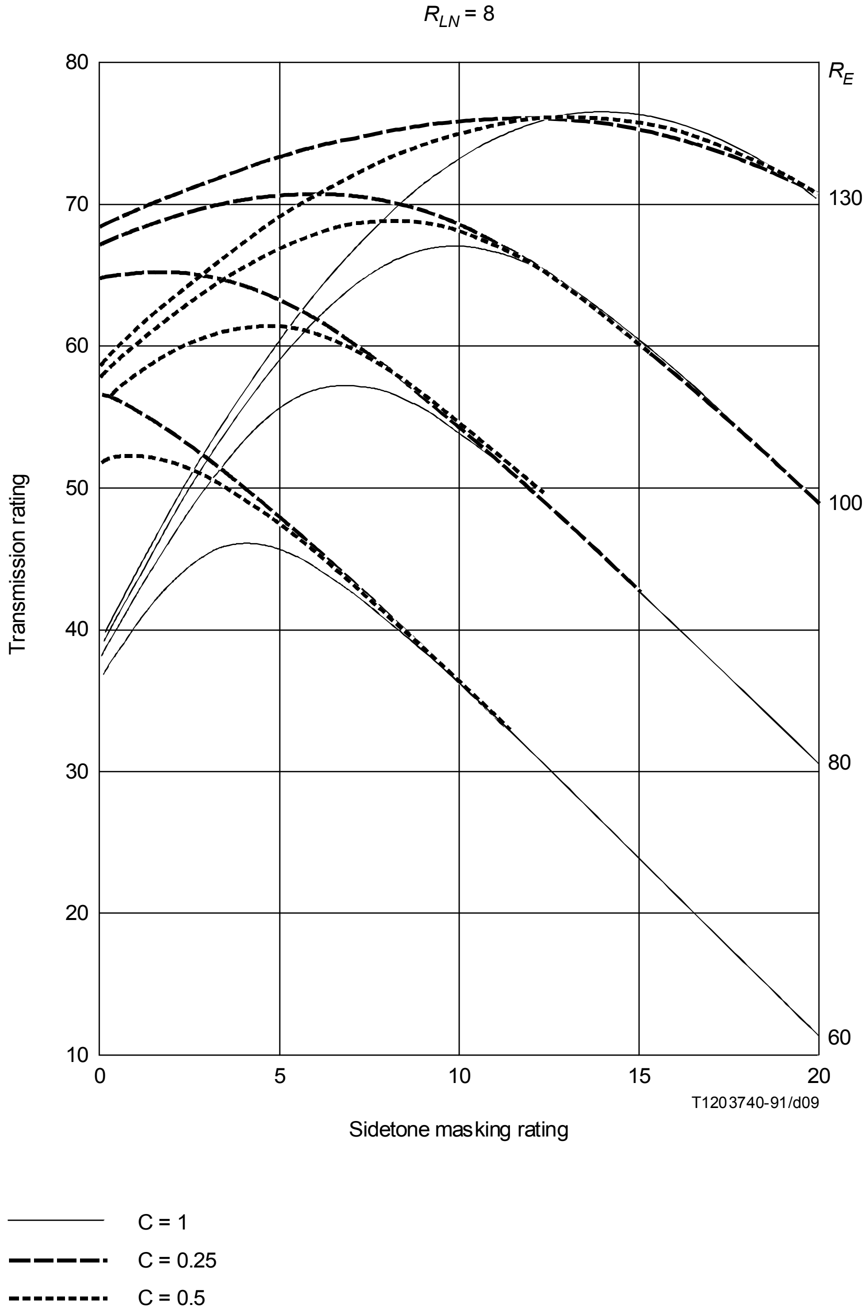


FIGURE 1-9

Transmission rating for OLR, circuit noise, talker echo and sidetone

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

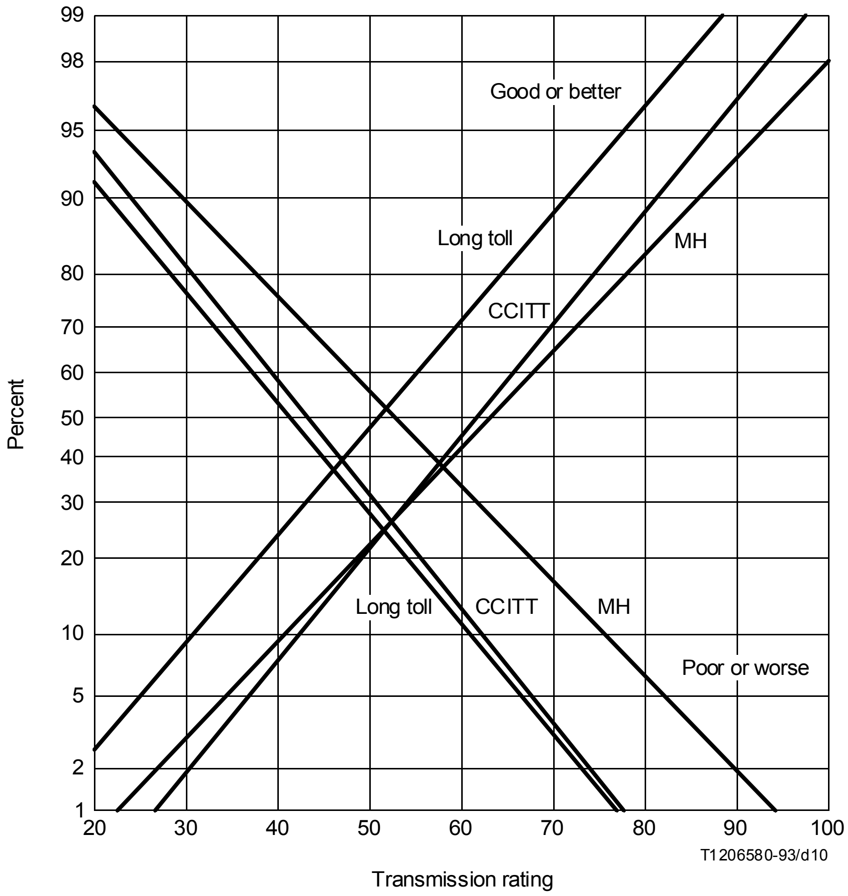


FIGURE 1-10

Comparison of opinion ratings as a function of transmission rating

Data Base[2]	*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

[2] 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.

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)[3)]

Summary

British Telecom makes extensive use of a theoretical model for predicting the transmission performance of telephone connections. A 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.

2.1 Types of model

Two types of "model" were recognized for predicting the performance of complete telephone connections in conversation. The first kind, exemplified in clause 1, 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 [13]) 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.

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.

The program CATPASS – 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 [14] and [15], but the

[3)] Formerly, Supplement No. 4, *Red Book*.

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 [16].

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 which (notwithstanding the statement made in the earlier version of this Supplement [17]) 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 Recommendation P.80. 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 Recommendation P.80. 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_{LE}, for each particular listening level and each circuit condition.

More detailed information about both conversation tests and listening tests may be found in [18], and also in Recommendation P.80.

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 [14];

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 [19].

From these data the loudness ratings are calculated. With speech level fixed, Y_{LE} 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 not given in detail here. The loudness part of the model is important in its own right, 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 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_{LE} 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

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

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

LOI is proportional to $\int B'_f P(Z_f)\,df$, 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

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

where

B'_i is the frequency weighting for the ith band (shown diagrammatically in Figure 2-1);

Z_i is the mean Z in the ith 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 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).

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_{LE} 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 $\mathrm{LOI}_{\mathrm{LIM}}$, introduced to cater for this effect, specifies the value of LOI that corresponds to maximum Y_{LE}; 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

$$\ln\left(\frac{Y_{LE}}{4 - Y_{LE}}\right) = 1.465\left[\ln\left(\frac{\mathrm{LOI}}{\mathrm{LOI}_{\mathrm{LIM}} - \mathrm{LOI}}\right) - 0.75\right]$$

as shown in Figure 2-5. This brings us to the point where Y_{LE} 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 [20].

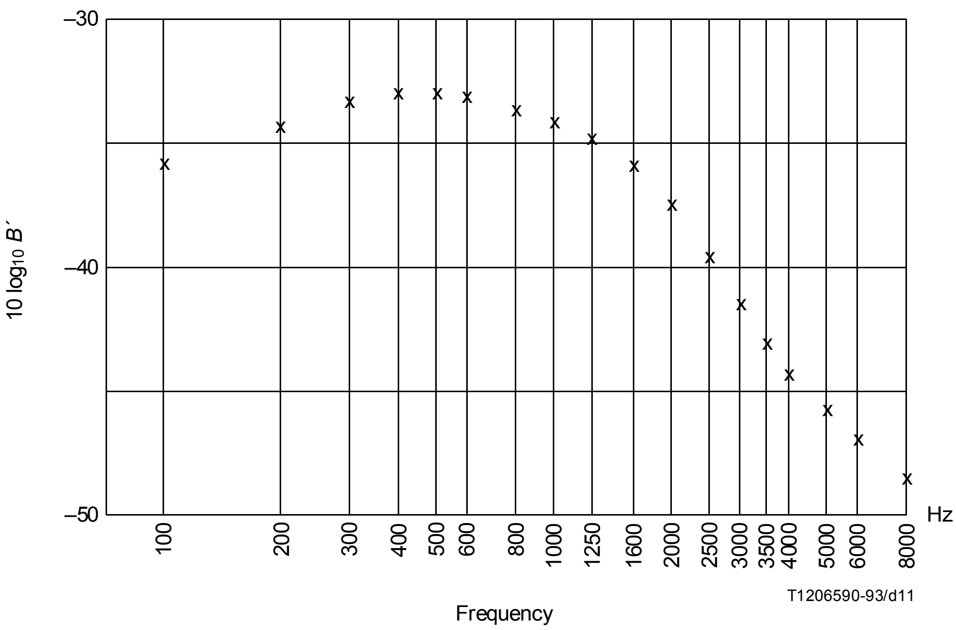


FIGURE 2-1

Frequency-weighting factor *B′* for Listening Opinion Index

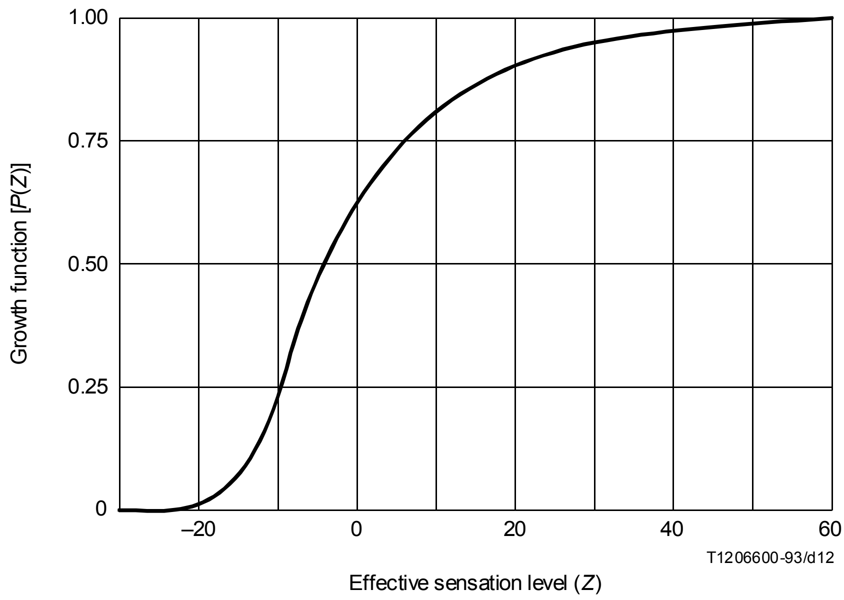


FIGURE 2-2

Growth function *P*(*Z*)

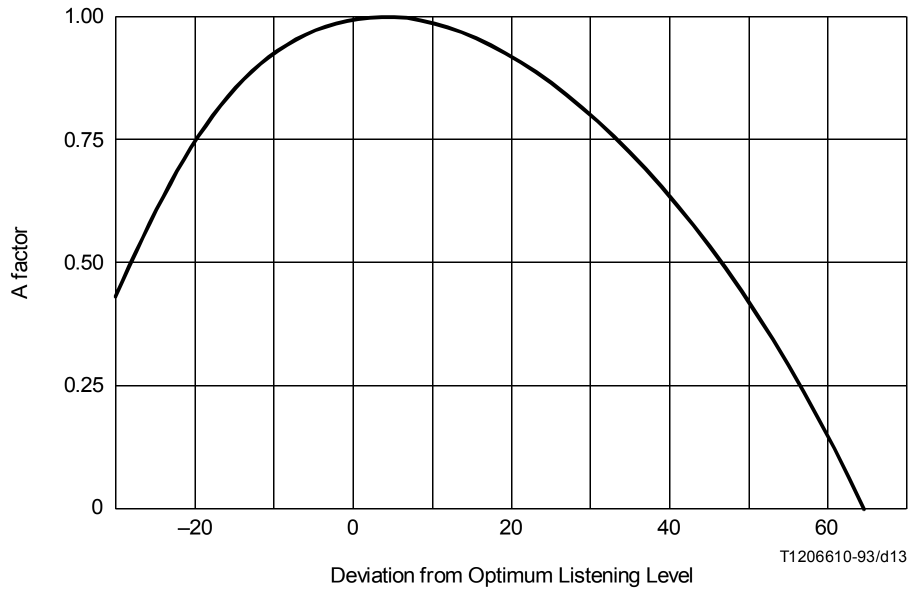


FIGURE 2-3

Effect of listening level on Listening Opinion Index

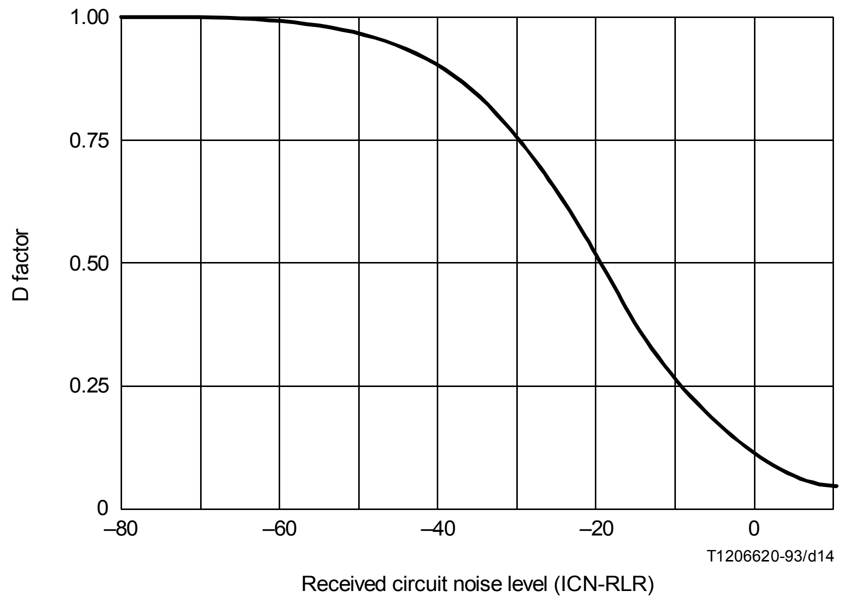


FIGURE 2-4

Effect of received noise level on Listening Opinion Index

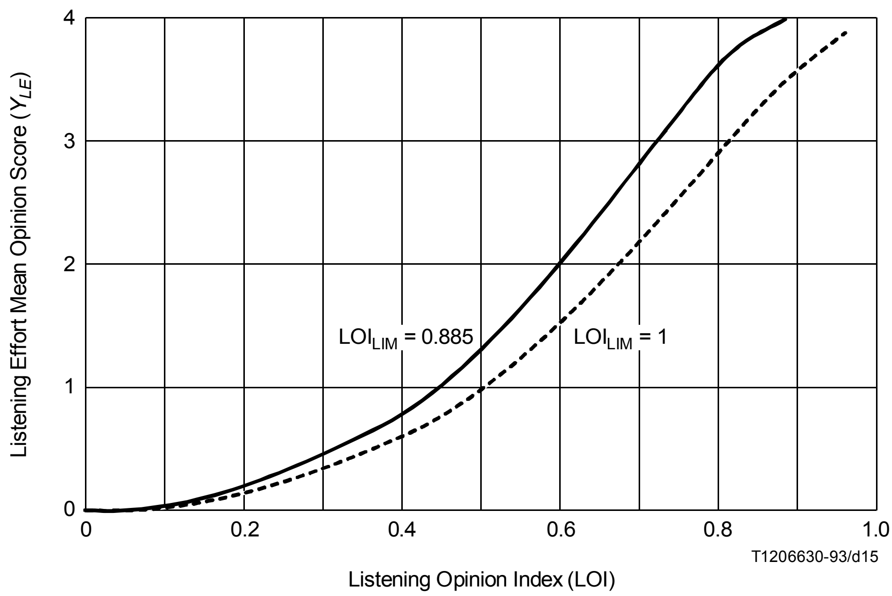


FIGURE 2-5

Listening Opinion Score as a function of Listening Opinion Index

2.7 Calculation of Conversation Opinion Score

In order to convert a value of Y_{LE} 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_{LE} at end *A*, which depends on V_C at end *B*, which depends on Y_{LE} at end *B*, which depends in turn on V_C at end *A*. Thus there is a circular dependence or feedback effect.

The sidetone paths introduce complications when STMR < 13 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.

$$\ln\left(\frac{Y'_C}{4 - Y'_C}\right) = 0.7\left[\ln\left(\frac{Y_{LE}}{4 - Y_{LE}}\right) + 0.5 - \frac{\mathrm{K}(13 - \mathrm{STMR})}{20}\left(\frac{4 - Y_{LE}}{Y_{LE}}\right)^2\right] \tag{2-1}$$

$$V_C - V_L = 4.0 - 2.1\, Y'_C - 0.3\,\mathrm{K}\,(13 - \mathrm{STRM}) \tag{2-2}$$

$$\ln\left(\frac{Y_C}{4 - Y_C}\right) = 0.8451 \ln\left(\frac{Y'_C}{4 - Y'_C}\right) - 0.2727 \tag{2-3}$$

where

K = 1 if STMR < 13

K = 0 otherwise

By substituting in equation (2-1) the value of Y_{LE} 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_{LE} 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_{LE} 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_{LE} in Figure 2-8.

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 (see Recommendation P.80).

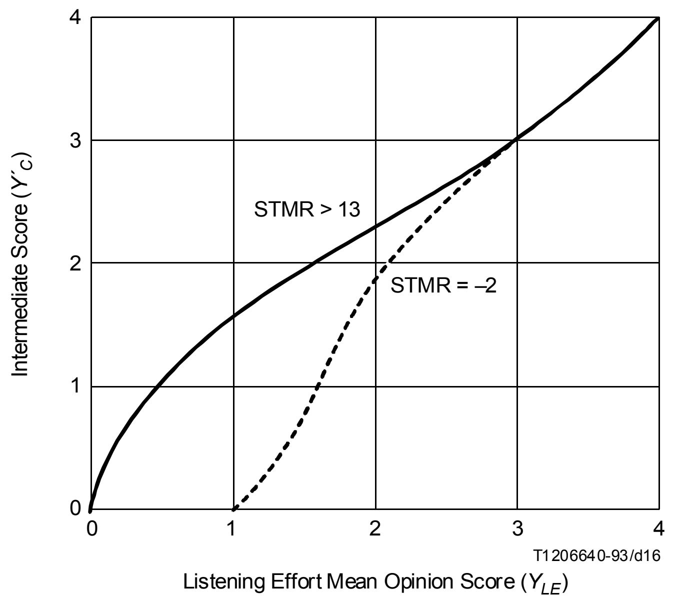


FIGURE 2-6

Intermediate Score as a function of Listening Opinion Score

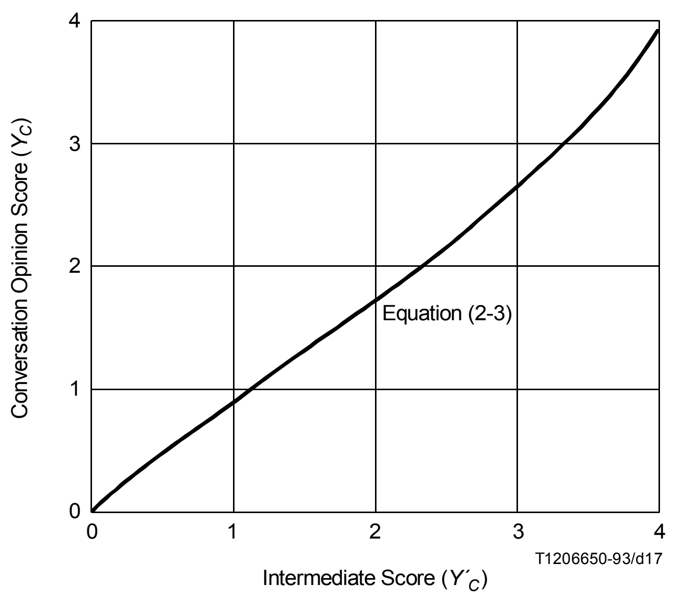


FIGURE 2-7

Conversation Opinion Score as a function of Intermediate Score

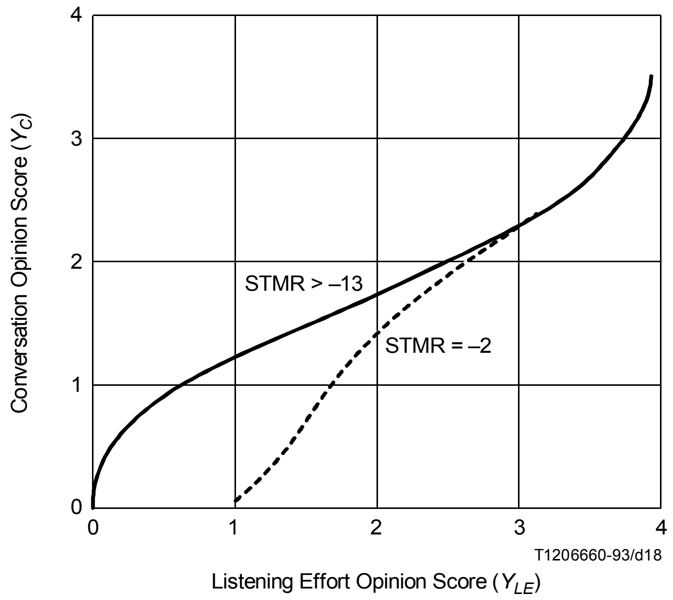


FIGURE 2-8

Conversation Opinion Score as a function of Listening Opinion Score

For example, percentage "Difficulty" is represented by the equation

$$\ln\left(\frac{D}{1 - D}\right) = -2.3 \ln\left(\frac{Y_C}{4 - Y_C}\right)$$

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 [23] 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.

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_C	Y_C	V_C	Y_C
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

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 [21] 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_{LE}. 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 > –5 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.

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	$N + S/2$	$2N$
Level at same point in absence of compander	S	N	N
Improvement	–	$-S/2$	$-N$

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)
= $-S/6 - 2N/3$

Hence

equivalent noise level = $N -$ improvement
= $N + S/6 + 2N/3 = S/6 + 5N/3$

This noise level is recalculated from V_C on each iteration and used to calculate the next value of Y_{LE}.

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_{LE} 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 [24]. This quantity can be calculated from the results of objective measurements and some fundamental data on speech and hearing. The theory takes into account transmission loss, circuit noise, room noise, attenuation/frequency distortion, sidetone and various distortions occuring in digital transmission. The effect of other types of non-linear distortion is under study.

The model used here belongs to the second type mentioned in [25] and in 2.1, since it reflects the cause-and-effect relationships between the input (properties of the connection considered, acoustic environment, some properties of speech and hearing) and the output (mutual information transmitted between speaker and listener). This clause only describes the practical method for performing the computation of the information index. As shown in [24] 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).

TABLE 3-1

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

NOTES
1 Over the transmitted band.
2 At frequency f_j.

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

$$Q_P = 10 \log_{10}\left[\sum s^2 / \sum (s - r)^2\right] \text{ dB} \tag{3-1}$$

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:

$$Q_{seg} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log_{10} \frac{\sum_{j=1}^{128} s^2 (j + 128\, m)}{\sum_{j=1}^{128} [s (j + 128\, m) - r (j + 128\, m)]^2} \text{ dB} \tag{3-2}$$

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 [24], is given by

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

with

$$V_j = \frac{3}{0.10 + 10^{-(Q_j + C_j)/10}} \tag{3-4}$$

B_j is the weight allocated to the band of rank j; $C_j = 10 \log_{10}(f_j/\Delta f_c)$, 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 60 and 70 of I-1, 70 and 80 of I.2. 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.

TABLE 3-2

Frequency weighting

j	Equal articulation bands Extreme frequencies (Hz)		$B_j \times 10^5$	C_j (dB)
1	200	330	5457	4.1
2	330	430	4733	5.6
3	430	560	6682	6.4
4	560	700	7497	6.9
5	700	840	6546	7.4
6	840	1000	6622	7.8
7	1000	1150	5585	8.0
8	1150	1310	5400	8.0
9	1310	1480	5273	8.2
10	1480	1660	5117	8.2
11	1660	1830	4517	8.2
12	1830	2020	4706	8.2
13	2020	2240	5073	8.2
14	2240	2500	5561	8.2
15	2500	2820	6310	8.2
16	2820	3200	6886	8.1
	TOTAL		102 158	

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 260-370 of I.2 is used. The law depends on a parameter $d = Q_{seg} - Q_p$. Numerical computations have shown that this law is valid both for PCM (d = 0) and for natural speech (d = −5.33) [24]. 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 programs used are reproduced in I-1 and I.2.

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 and the mean opinion scores (MOS) for listening determined in the CNET Laboratory [26].

Table 3-4 shows the correlation of these MOS with the information index (see 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 attenuator. 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[4)] (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 kHz to 8 kHz.

3.3.2.1 Basic data

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

BK is the hearing threshold for continuous-spectrum sounds (L_s in [24]) referred to ear reference point;

S is the 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;

BJ is the frequency weighting (see [24]);

CJ is the correction term in formula 3-4 giving V_j.

3.3.2.2 Electroacoustic characteristics

These depend on the connection considered and are defined in III.1.2; here it is sufficient to know that RLR (receiving loudness rating) = −3.9 dB.

NOTE – SLR (for sending) is not needed in principle, but may be used for a small correction to OLR if a different item of the same type of telephone set is used.

The overall loss OLA between mouth (MRP) and ear (ERP) reference points, as would be measured with an artificial ear, was calculated as shown in III.2.1.

4) See Table 3-5.

TABLE 3-3

Examples of measured values of Q (dB) and calculated I_I

System	MOS (0-4)	Over total band		Q_s in band No.																I_t (dB)
		Q_P	Q_{seg}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
MNRU 15 dB	1.19	15	14.9	20.1	17.9	15.3	11.5	8.1	5.5	3.3	1.7	2.2	1.9	0.9	0.5	0.3	0.6	0.4	€–€0.6	17.1
20	1.84	20	19.8	25.1	23	20.3	16.6	13.2	10.6	8.3	6.8	7.2	6.9	5.9	5.4	5.3	5.6	5.4	4.4	22.5
25	2.44	25	24.9	30.4	28.5	25.6	21.3	17.5	15	12.6	10.8	11.4	11.1	10.1	9.6	9.5	9.8	9.4	8.2	25.2
30	2.81	30	29.9	35.6	33.5	30.4	26.2	22.2	19.6	17.3	15.5	16.1	15.9	14.9	14.4	14.2	14.6	14.2	13	26.7
MNRUS 15	2.22	15.1	15.4	13.6	13.8	13.1	10.8	8.8	7.7	6.6	6	7.7	8.2	7.9	8.2	8.7	9.7	10.1	9.7	22.8
20	2.59	20.1	20.4	18.6	18.9	18.1	15.7	13.6	12.5	11.5	10.8	12.5	13.1	12.8	13	13.6	14.7	15	14.6	25.8
25	3.25	25.1	25.4	23.9	24.2	23.2	20.8	18.6	17.3	16.4	15.7	17.3	17.9	17.7	17.9	18.4	19.4	19.8	19.1	27.0
30	3.31	30.1	30.4	29.1	29.3	28.2	25.8	23.4	22.2	21.2	20.5	22.1	22.8	22.5	22.7	23.1	24.1	24.5	23.9	27.4
F 16 kbit/s	0.81	14.2	11.4	15.2	13.5	10.8	7.3	4.1	1.7	−0.9	−2.1	−1.5	−1.9	−2.8	−3.3	−3.2	−2.5	−2.5	−2.7	16.5
24	1.59	20	17.7	21.7	19.8	17.4	13.9	10.6	8.1	5.9	4.7	5	4.5	3.8	3.5	3.1	3.6	3.4	2.8	23.4
32	2.59	25.2	22.8	27.1	25.4	22.6	19.2	16	13.7	11.4	10	10.3	9.8	9	8.6	8.4	8.8	8.5	7.9	26.0
V 16 kbit/s	2.03	12.8	14.5	16.2	14.2	11.9	8.6	6.1	4.2	2.5	1.9	2.6	2.2	2.1	2	2.4	2.8	2.9	2.8	24.1
24	3.06	19.7	21.8	25.1	23.2	20.6	17	14.1	11.8	9.3	8.2	8.8	8.3	7.3	7.1	7	7.7	7.7	7.2	26.7
32	3.28	25.7	28.3	32.1	30.1	27.6	24	20.5	18.3	15.6	14.4	14.8	14.3	13.3	13.1	12.8	13.3	13.2	12.3	27.4
SB 24 kbit/s	2.84	17.1	15	16.7	14.7	13	16.7	15.3	13.4	12.9	14.8	14.9	12.4	11.7	11.3	11.3	11.6	7.4	6.3	26.2
32	3.22	20.4	19.4	22.1	20	18.4	21.9	20.3	18.5	17.1	19.6	20	20.2	19.9	17.3	16.7	16.4	12.7	11.4	27.2

MNRUS MNRU with shaped noise
F ADPCM with fixed predictor
V ADPCM with variable predictor
SB Sub-band coding

TABLE 3-4

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

Objective measure (Note)	System									
	Group B PCM, ADM, ADPCM-F		Group A same as B + ATC APC – AB		Group F ADPCM-F, ADPCM-V		Group E same as F + sub-band coding		Group D same as E + MNRU shaped MNRU	
	R	S	R	S	R	S	R	S	R	S
Q_P (SNR)	0.798	0.578	0.803	0.559	0.687	0.680	0.590	0.711	0.650	
Q_{seg} (SNRseg)	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						
Cepstral distance	0.954	0.208	0.929	0.331						
SNRF									0.884	
Information index					0.992	0.104	0.993	0.109	0.976	0.174

SNRF Q_s frequency weighted
R Correlation coefficient
S Standard deviation (in terms of MOS on a 0 to 4 or 1 to 5 scale)
NOTE – Notations from Table 3-1.

3.3.2.3 Noise components

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

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

BDCN is the spectrum of circuit noise;

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

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

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

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

CN is the BDCN computed for $ICN_0 = -60$ dBmp;

ST is the BDST computed for RN = 50 dBA and STMR = 15 dB;

EL is the BDEL computed for RN = 50 dBA.

The computations were made from the frequency characteristics given in Tables 4-4, C.1 and D.1. The circuit noise spectrum used in the example of program of Appendix II corresponds to the last column of Table C.1 (see III.1.1) and applies to almost all cases in Tables 3-5 and 3-6. A different spectrum is used for white noise.

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 720-820) which is similar to the formulas of Annex A/P.11 but was adapted to the results of subjective opinion tests published in [28].

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 [24] clause V.2) 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 [29], 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 [24] clause V.3. 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 [28], 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.

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)	ICN (dBmp)	ICN_0 (dBmp)	OLR (dB)	Y_L	I_L (dB)	Y_C	I_C (dB)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
60	–62.1	–58.2	1.4	3.13	22.37	2.94	22.2
		(exchange	11.4	2.5	21.26	2.34	21.11
		noise	21.4	2.31	17.32	1.58	17.87
		only)	31.4	0.65	8.16	0.2	13.64
60	–59.8	–55.9	1.4	3.1	22.36		
			11.4	2.91	21.15		
			21.4	1.75	16.92		
			31.4	0.8	7.37		
60	–55.8	–51.9	1.4	2.83	22.38	2.99	22.2
			11.4	2.75	20.97	2.39	20.83
			21.4	1.79	16.34	1.28	17.24
			31.4	0.5	6.51	0.43	12.4
60	–51.4	–47.5	1.4	3.06	22.37	3.08	22.17
			11.4	2.24	20.53	2.17	20.42
			21.4	1.05	14.97	1.29	16.41
			31.4	0.09	5.17	0.22	11.11
60	–45.6	–41.7	1.4	2.31	22.18	2.63	21.95
			11.4	1.4	19.22	1.73	19.34
			21.4	0.64	11.66	0.77	14.77
			31.4	0.05	3.15	0.13	8.55

Explanation of columns

(1) Room noise, dBA

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

(3) $ICN_0 = ICN + 3.9$ dB

(4) OLR (see Recommendation P.79)

(5) Listening MOS (on a 0 to 4 scale), test 2 of [28], p. 4-4

(6) Listening information index (position L of Appendix II)

(7) Conversation MOS, test 6 of [28], p. 4-9

(8) Conversation opinion index (position C of Appendix II)

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)	ICN (dBmp)	ICN_0 (dBmp)	OLR (dB)	Y_L	I_L (dB)
0		−100	−3.6	2.30	21.05
			1.4	2.83	21.86
			6.4	3.26	22.51
	(Note 1)	(Note 1)	11.4	2.92	22.16
			16.4	2.59	21.43
			21.4	2.12	20.48
			26.4	1.89	19.24
			31.4	1.23	17.07
60	−55.8	−51.9	−3.6	2.61	21.35
			1.4	2.94	22.38
			6.4	3.00	22.15
			11.4	2.38	20.97
			16.4	1.80	19.09
			21.4	1.41	16.34
			26.4	0.91	11.95
			31.4	0.44	6.51
60	−56.9	−53	1.4	3.20	22.39
		(Note 2)	11.4	2.53	21.06
			21.4	1.24	16.62
			31.4	0.24	6.77
50	−55.8	−51.9	1.4	3.21	22.67
			11.4	2.64	21.7
			21.4	1.58	18.48
			31.4	0.35	11.16
45	−64.9	−61	1.4	3.23	22.52
			13.4	2.62	21.77

NOTES

1 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.

2 White noise.

TABLE 3-7

Correlation between MOS and the information index in the case of analogue transmission[a)]

Type of connection	Range of conditions (Note)			Type of MOS	Model	Correlation coefficient	Deviation in terms of Y	
	RN (dBA)	ICN_0 (dBmp)	OLR (dB)				Standard error	Extreme deviations
NTT 600-type telephone sets + SRAEN filter	0 to 60	$-\infty$ and -61 to -51.9	-3.6 to $+31.4$	Y_L	I	0.985	0.11	$-0.27 + 0.22$
					O		0.17	$-0.37 + 0.32$
	60	-58.2 to -41.7	$+1.4$ to $+31.4$	Y_C	I	0.977	0.16	$0.31 + 0.32$
					O		0.20	$-0.38 + 0.36$

a) This table replaces Table VI of [24].

I Information index

O OPINE model

NOTE – Notations are of Appendix II.

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 procedure is explained in Appendix III.

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

3.4.2 Connection including digital processes

Subclause 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 [24].

Anyhow, when digital processes 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.

4 Overall Performance Index model for Network Evaluation (OPINE) (Contribution by NTT)

4.1 Introduction

NTT has been studying an objective model for evaluating telephone transmission performance [30], [31], [32], [33]. This describes OPINE (Overall Performance Index model for Network Evaluation), focussing on practical use.

OPINE deals with transmission loss, circuit noise, room noise, attenuation/frequency distortion (fundamental factors), quantizing distortion, talker echo and sidetone. It models the auditory-psychological process of evaluation by human beings of telephone transmission performance based on these factors. It is therefore the second type of model according to the classification of clause 2 (British Telecom). The model's basic principle is the fact that evaluation of psychological factors (not physical factors) on the psychological scale is additive. The model is extended from the first revision to take additional physical factors into account.

OPINE was first constructed for fundamental factors in 1983 [30]. The opinion test data used for coefficient training and verification largely depend on the results of the experiment conducted at NTT ECL, Musashino in 1975. Its main purpose was to study the opinion score as a speech quality measure and a basis of telephone transmission standard. [31] describes the raw data. The experiment was of large scale with various factors taken into account, using an NTT 600-type telephone set.

In 1985, opinion tests were conducted for quantizing distortion. A newer revision of the model that also dealt with quantizing distortion was formulated and verified [32].

Some further opinion tests for talker echo and sidetone were conducted in parallel [35], [36]. A study of the evaluation characteristics of talker echo and its interaction with loudness was undertaken later.

In 1986 revision 2.0 of OPINE was formulated in which all the parameters were rewritten in terms of loudness rating (LR). This revision was improved and updated to 2.1. Improved points in revision 2.1 are these minor changes:

- Δf has been corrected to agree with that of Recommendation P.79;
- a trivial bug of the Fortran program in revision 2.0 has been eliminated.

While the model configuration was studied, the psychological characteristics of opinion evaluation were also investigated [36], using transmission loss and circuit noise as variables. The main conclusions were

- the opinion score has good reproducibility if experimental design, subject type and other conditions are kept constant;
- the test condition range greatly affects the opinion score. The loss condition range especially affects the absolute opinion score.

In spite of the above conclusions, an absolute evaluation for a given network condition needs to be defined for practical use.

Therefore, we specify two classes of opinion tests:

- Class 1, in which the score reflects the mean value of network evaluation for general telephone customers;
- Class 2, which produces a relative score but is sensitive to a few given physical factors.

In the class 1 test, the purpose is to obtain an absolute opinion score. Therefore the range of test conditions should be similar to that for degradation in the present commercial network. The more factors taken into account in the opinion test, the closer the score comes to an absolute value. The number of subjects should exceed 60. The class 2 test, on the other hand, is used to study interaction among several factors. It is more practical but the score obtained is not absolute. For this test, it is desirable that the subject's occupation be connected with the subject of speech quality.

In formulating OPINE, we classified the opinion database in 1975 as the first class, and the rest as the second.

Opinion data executed after 1983 were mainly used for qualitative verification of the additive characteristics of evaluation on a psychological scale for different factors.

In extensions of OPINE, coefficients for newer factors were changed so that they fitted the results of the absolute score of the class 1 test of 1975.

4.2 Outline of the model

Five psychological factors affecting telephone speech quality were chosen on the basis of previous studies:

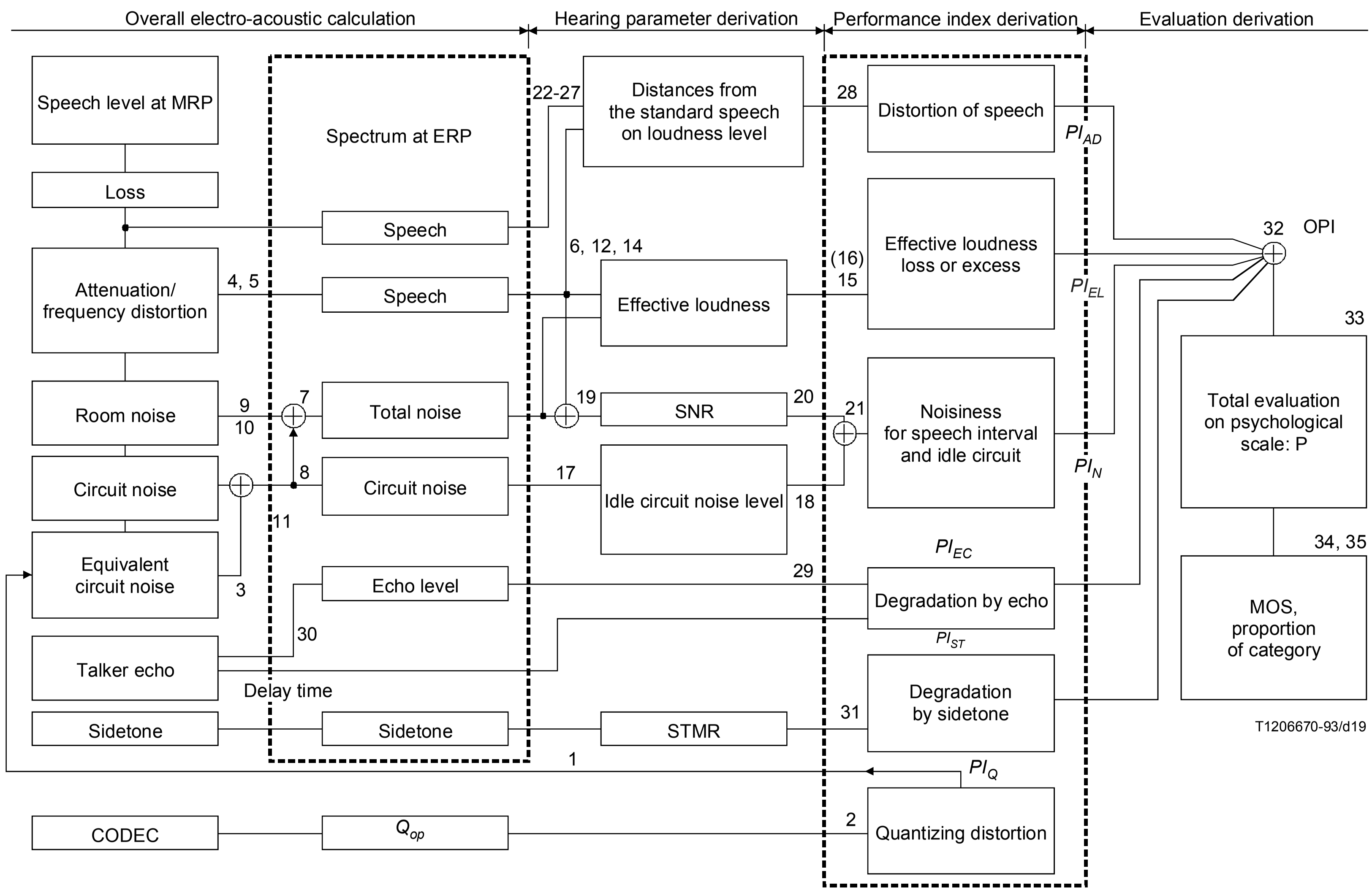
1) speech distortion for attenuation/frequency distortion;
2) effective loudness loss or excess in speech;
3) noisiness during speech intervals and non-speech intervals;
4) degradation caused by talker echo;
5) degradation caused by sidetone.

A PI (Performance Index) is also introduced for each of the above factors which indicates the psychological degradation degree. The MOS is estimated from the Overall Performance Index (OPI) which is obtained by summing up all PIs.

To calculate the PI for each factor, physical factors are obtained for loudness, distortion, etc., and each PI is transformed by an appropriate function. These functions are determined heuristically and the necessary constants are estimated from subjective data. The degree to which each factor influences the evaluation is reflected by these constants. The conceptual block diagram of OPINE is shown in Figure 4-1. The model consists of four parts:

1) an overall electro-acoustic calculation;
2) hearing parameter derivation;
3) a performance index derivation; and
4) an evaluation derivation.

The numbers in the figure refer to the equation numbers listed in 4.3.



MRP Mouth reference point
ERP Ear reference point
Q_{op} Opinion equivalent speech to speech correlated noise ratio

NOTE – The numbers in the figure refer to equation numbers in 4.3.

FIGURE 4-1

Block diagram of OPINE

4.3 Configuration of OPINE

All the symbols are classified into 5 types:

Type [A]: Model parameters

Type [A-1]: Constants or coefficients adopted from standards

Type [A-2]: Constants or coefficients that OPINE accepted from results of other studies

Type [A-3]: Estimated coefficients from the results of NTT's subjective tests

Type [B]: Input variables of the section being described

Type [C]: OPINE's intermediate outputs of the section being described.

Input variables to the model and the values of model parameters are listed in 4.4. In the following equations, C_j ($j = 1,13$) denote constants ([A-3]-type). The suffix i denotes the 1/3 octave frequency band number. Relations among variables corresponding to each section are shown in Figures 4-3 through 4-10. The definition of the graphic symbols used in these figures is shown in Figure 4-2.

4.3.1 Overall electro-acoustic calculation

4.3.1.1 Opinion equivalent white noise level of quantizing distortion

The model expresses CODEC's subjective evaluation as an opinion equivalent speech-to-speech correlated noise (Q_{op}). Then the equivalent white noise level is acquired using the subjective opinion test results for MNR. If A_{op} of a certain CODEC or its tandem connection is known, it is possible to use the value as input. The various CODECs and Q_{op} adopted here are listed in Table 4-1.

$$PI_Q = -0.0000218\, Q_{op}^3 + 0.00489\, Q_{op}^2 - 0.283\, Q_{op} + 4.915 \tag{4-1}$$

$$V_{Wop} = -2.022\, PI_Q^3 - 7.51\, PI_Q^2 + 21.9\, PI_Q - 76.9 - (OLR - 6.4) - (RLR + 3.8) \tag{4-2}$$

$$V_{CQ} = V_C\,(+)\,V_{Wop} \tag{4-3}$$

TABLE 4-1

Values of Q_{op} for PCM and ADPCM$_v$

Transmission system	Q_{op}
PCM µ-255, 8 bit	36.0
7	32.8
6	27.7
5	22.5
4	16.7
ADPCM$_v$	29.2

where

(+) is the power summation operation.

Type [B] symbols

Q_{op} is the opinion equivalent speech-to-speech correlated noise ratio (dB);

V_C is the circuit noise level at the input to the receiving local telephone circuit (dBmp);

OLR is the overall loudness rating of the telephone system being considered (dB);

RLR is the receive loudness rating of the telephone system being considered (dB).

Type [C] symbols

V_{Wop} is the opinion (PI) equivalent white noise level at the input to the receiving local telephone circuit (dBmp);

PI_Q is the PI for quantizing distortion;

V_{CQ} is the equivalent circuit noise level when both circuit noise and quantizing distortion are present (dBmp).

NOTE – When the digital system is not considered in a test condition, equations (4-1) and (4-2) are not necessary, and V_{Wop} is set to an arbitrary low level, such as −100, in equation (4-3).

4.3.1.2 Speech level and total noise level at an ERP (see also Annex C)

$$S_i = B_{Si} - L_{MEi} + 10 \log_{10} \Delta f_i \tag{4-4}$$

$$S_{Pi} = B_{Pi} - L_{MEi} \tag{4-5}$$

$$X_i = B_{0i} - K_i \tag{4-6}$$

$$N_i = N_{CQi} \, (+) \, N_{RNSTi} \, (+) \, N_{RNEi} + 10 \log_{10} \Delta f_i \tag{4-7}$$

$$N_{CQi} = V_{CQi} + S_{JEi} \tag{4-8}$$

$$N_{RNSTi} = B_{RNi} + L_{RNSTi} \tag{4-9}$$

$$N_{RNEi} = B_{RNi} + L_{RNEi} \tag{4-10}$$

$$N'_{CQi} = N_{CQi} + 10 \log_{10} \Delta f_i \tag{4-11}$$

where

(+) is the power summation operation.

Type [A-1] symbols

B_{Si} is the spectrum density of speech referred to an MRP (dB rel 20 µPa/Hz);

Δf_i is the width of ISO preferred 1/3 octave frequency band (Hz).

Type [A-2] symbols

B_{Pi} is the peak spectrum level of speech referred to an MRP (dB rel 20 µpa/Hz);

X_i is the hearing threshold for the continuous sound referred to an ERP (dB rel 20 µPa/Hz);

B_{0i} is the pure tone audibility threshold (dB rel 20 µPa/Hz);

K_i is the critical bandwidth (dB);

L_{RNEi} is the leakage transmission loss at a listener's ERP (dB).

Type [B] symbols

L_{MEi} is the overall mouth-to-ear loss (dB);

S_{JEi} is the receiving sensitivity of a local telephone circuit from the electrical input to an ERP (dB rel Pa/V);

B_{RNi} is the room noise spectrum density (dB rel 20 μPa/Hz).

A-weighted evaluation of B_{RNi} becomes R_N (dBA).

L_{RNSTi} is the sidetone transmission loss from an MRP to an ERP (dB);

V_{CQi} is the equivalent circuit noise level when both circuit noise and quantizing distortion are present (dBV/Hz).

Psophometric weighted evaluation of V_{CQi} becomes V_{CQ}.

Type [C] symbols

S_i is the band spectrum level of speech at an ERP (dB rel 20 μPa/Hz);

S_{Pi} is the peak spectrum level of speech referred to an ERP (dB rel 20 μPa/Hz);

N_i is the total band noise level at an ERP (dB rel 20 μPa);

N_{CQi} is the noise level caused by stationary circuit noise and quantizing distortion at an ERP (dB rel 20 μPa/Hz);

N'_{CQi} is the band level of N_{CQi} (dB rel 20 μPa);

N_{RNSTi} is the noise sidetone level caused by room noise at an ERP (dB rel 20 μPa/Hz);

N_{RNEi} is the room noise level via earcap leakage (dB rel 20 μPa/Hz).

4.3.2 Derivation of hearing parameters and performance index (PI)

4.3.2.1 PI_{EL} (*PI* for effective loudness loss or excess)

$$\lambda_E = C \sum_{i=1}^{M} 10^{\frac{-m(L_{MEi} + b_n)}{10}} G_i \Delta f_i \quad \text{(from Recommendation P.79)} \tag{4-12}$$

$$b_n = 44{,}38 \exp(-0{,}0869\, e_n) \tag{4-13}$$

$$e_n = \left[S_{Pi} - \left\{X_i (+)(N_i - 10\log_{10})\Delta f_i)\right\}\right]_{\max} \tag{4-14}$$

$$PI_{EL} = \sqrt{\frac{C_2}{C_1}\left(10 \log_{10} \frac{\lambda_E}{\lambda_0}\right)^2 + C_2} - \sqrt{C_2} \tag{4-15}$$

where

max is a suffix which denotes maximum value within the passing bands.

Type [A-1] symbols

G_i is the ratio of loudness for frequency band *i* in a lossless system to total loudness (loudness function);

Δf_i is the width of the *i*th frequency band (Hz);

m is the ear's exponential coefficient (= 0.175);

M is the number of partitioned bands (= 19).

Type [A-3] symbols

λ_0 is the optimum loudness at ERP;

C is a constant. Value of C is not needed since C is cancelled in equation (4-15).

Type [B] symbol

L_{MEi} is the transmission loss-frequency characteristic from MRP to ERP (dB).

Type [C] symbols

PI_{EL} PI on loudness in both the absence and presence of noise;

λ_E is the effective loudness at ERP taking the effect of noise into account;

b_n is the equivalent loudness loss in the presence of noise (dB);

e_n is the maximum sensation peak level of speech (dB).

4.3.2.2 Expression of PI_{EL} in terms of loudness rating (LR)

Equation (4-15) is theoretically expressed in terms of LR. The derivation of equation (4-16) from equation (4-15) is shown in Annex E.

$$PI_{EL} = \sqrt{\frac{C_2}{C_1} m^2 (OLR + b_n - OLR_0)^2 + C_2} - \sqrt{C_2} \tag{4-16}$$

where

Type [A-3] symbol

OLR_0 is the overall loudness rating value at which the telephone system supplies the optimum loudness (dB).

Type [B] symbol

OLR overall loudness rating of the telephone system being considered (dB).

4.3.2.3 PI_N (PI for noisiness)

$$N'_i = \begin{cases} N'_{CQi} - N_{th} & \text{when } N'_{CQi} \geq N_{th} \\ 0 & \text{when } N'_{CQi} < N_{th} \end{cases} \tag{4-17}$$

$$PI_{IN} = C_3 \sum_{i=1}^{M} \left\{ 10^{\frac{A_i}{10}} \left(10^{\frac{nN'_i}{10}} - 1 \right) \right\} \tag{4-18}$$

where

Type [A-1] symbol

A_i is the weight for A-characteristic at frequency band i (dB).

Type [A-3] symbols

N_{th} is the noise threshold (dB rel 20 µPa);

n is the exponent.

Type [B] symbol

N'_{CQi} (see 4.3.1.2).

Type [C] symbols

PI_{IN} is the PI for idle circuit (non-speech interval) noisiness;

N'_i is the level above the noise threshold (dB).

$$SNR = 10\log_{10}\left(\sum_{i=1}^{M} 10^{\frac{Si}{10}} \Big/ \sum_{i=1}^{M} 10^{\frac{Ni}{10}}\right) \tag{4-19}$$

$$PI_{SN} = \begin{cases} C_4(SNR - SNR_{th}) & \text{when } SNR \le SNR_{th} \\ 0 & \text{when } SNR > SNR_{th} \end{cases} \tag{4-20}$$

$$PI_N = PI_{IN} + PI_{SN} \tag{4-21}$$

Type [A-3] symbol

SNR_{th} is the threshold below which the signal-to-noise ratio has no effect on the evaluation (dB).

Type [B] symbols

S_i (see 4.3.1.2)

N_i (see 4.3.1.2)

Type [C] symbols

PI_{SN} is the *PI* for speech interval noisiness;

SNR is the signal-to-noise ratio at an ERP (dB).

4.3.2.4 PI_{AD} (*PI* for attenuation/frequency distortion)

$$D_1 = \sqrt{\frac{1}{M_s} \sum_{i=1}^{M_s} \Lambda_i^2} \tag{4-22}$$

$$D_u = \sqrt{\frac{1}{M - M_s} \sum_{i=M_s+1}^{M} \Lambda_i^2} \tag{4.23}$$

$$\Lambda_i = \begin{cases} \Lambda_{li} - \Lambda_{di} & \text{when } \Lambda_{li} - \Lambda_{di} \le \Lambda_{th} \\ \Lambda_{th} & \text{when } \Lambda_i > \Lambda_{th} \end{cases} \tag{4-24}$$

$$\Lambda_{li} = g_i\,(S_i + d_i) \tag{4-25}$$

$$\Lambda_{di} = g_i\,(S_i) \tag{4-26}$$

$$g_i(x_i) = \begin{cases} a_i + b_i x_i + c_i x_i^2 & \text{when } a_i + b_i x_i + c_i x_i^2 \ge L_{th} \\ L_{th} & \text{when } a_i + b_i x_i + c_i x_i^2 \ge L_{th} \end{cases} \tag{4-27}$$

$$PI_{BL} = C_5 D_1 + C_6 D_u \tag{4-28}$$

where

g_i is the conversion function from the speech power spectrum into a loudness level by equal-loudness curve (from [38]);

x_i is the arbitrary band speech level (dB rel 20 µPa).

Type [A-1] symbols

M is the number of partitioned bands (= 19);

a_i, b_i, c_i are the parameters for converting to loudness level (in phones); they are a function of frequency.

Type [A-2] symbol

M_s is the band number in which 1 kHz is contained (= 11).

Type [A-3] symbols

L_{th} is the loudness threshold (phon);

Λ_{th} is the threshold of Λi (phon).

Type [B] symbol

d_i is the relative loss caused by attenuation/frequency distortion between junctions (dB).

It is 0 dB at 800 Hz. $S + d$ represents hypothetical band speech level at an ERP without attenuation/frequency distortion (reference speech).

Type [C] symbols

Λ_i is the difference between reference speech and distorted speech (phon);

Λ_l is the loudness level converted from reference speech (phon);

Λ_d is the loudness level converted from speech with both loss and band limitation (phon);

D_u is the distance between Λ_l and Λ_d above 1 kHz;

D_l is the distance between Λ_l and Λ_d below 1 kHz;

PI_{AD} is the *PI* for attenuation/frequency distortion.

4.3.2.5 PI_{EC} (*PI* for talker echo)

$$PI_{EC} = \sqrt{\frac{C_8}{C_7}(-E + E_0)^2 + C_8} + \sqrt{\frac{C_8}{C_7}}(-E + E_0) \qquad (4\text{-}29)$$

$$E_0 \begin{cases} C_9 \log_{10} D + C_{10} & \text{when } 0 < D < 60 \\ C_{11} \log_{10} D + C_{12} & \text{when } D \geq 60 \end{cases} \qquad (4\text{-}30)$$

where

Type [B] symbols

E is the talker echo LR (dB);

D is the delay time of talker echo (msec).

Type [C] symbols

PI_{EC} is the performance index on talker echo;

E_0 is the critical talker echo LR (dB).

4.3.2.6 PI_{ST} (*PI* for sidetone)

$$PI_{ST} = \sqrt{\frac{C_{13}}{C_7} (-St + St_0)^2 + C_{13}} + \sqrt{\frac{C_{13}}{C_7}} (-St + St_0) \tag{4-31}$$

where

Type [A-3] symbol

St_0 is the critical STMR (dB).

Type [B] symbol

St is the STMR (sidetone masking rating) (dB).

Type [C] symbol

PI_{ST} is the performance index on sidetone.

4.3.3 Evaluation derivation (see also Annex D)

$$OPI = PI_{EL} + PI_N + PI_{AD} + PI_{EC} + PI_{ST} \tag{4-32}$$

$$P = P_0 - OPI \tag{4-33}$$

where

Type [A-3] symbol

P_0 is P with no degradation.

Type [C] symbols

OPI is the overall performance index;

P is the mean overall evaluation on this psychological scale.

$$MOS = \sum_{k=0}^{4} k\, p_k \tag{4-34}$$

or in practical form:

$$MOS = 4 - \sum_{k=0}^{3} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(K+0.5-P)/\sigma} \exp(-t^2/2)\mathrm{d}t \tag{4-35}$$

where

Type [A-3] symbol

σ is the standard deviation of normal distribution of P and OPI.

Type [C] symbols

MOS is the mean opinion score ranging from 0 to 4;

p_k is the ratio of evaluation category k to all the categories.

Equation (4-35) is calculated using the standard normal distribution table. The derivation of this equation from equation (4-34) is shown in Annex F.

Equations (4-34) and (4-35) are the adaptation of the model in [39].

4.4 Symbol types and values

Input variables to the model are listed in Table 4-2. L_{ME} and STMR can be calculated in advance using the method described in Recommendation P.79.

Values of a_i,b_i and c_i ([A-1]-type) are shown in Table 4-3. Values of other model parameters ([A-1]- and [A-2]-type parameters) are shown in Table 4-4. Values of estimated constants or coefficients from the subjective test results ([A-3]-type parameters) are shown in Table 4-5.

TABLE 4-2

Input variables to the model

Symbols	Definition
V_C	See 4.3.1.1
Q_{op}	See 4.3.1.1
OLR	See 4.3.1.1 and 4.3.2.2
RLR	See 4.3.1.1
S_{MJi}	Mouth to junction loss (dB rel V/Pa)
S_{JEi}	See 4.3.1.2
L	Junction to junction loss at 800 Hz (dB)
d_i	See 4.3.2.4
L_{MEi}	See 4.3.1.2
R_N	See 4.3.1.2
L_{RNSTi}	See 4.3.1.2
E	See 4.3.2.5
D	See 4.3.2.5
L_{MESTi}	Mouth to ear sidetone loss (dB)
St	See 4.3.2.6

NOTES

1 $L_{MEi} = -S_{MJi} - S_{JEi} + (L + d_i)$.

2 St is calculated according to 8//P.79.

3 S_{MJi}, L and L_{MEST} only necessary to calculate L_{MEi} and St.

4 R_N should be expanded B_{RNi}.

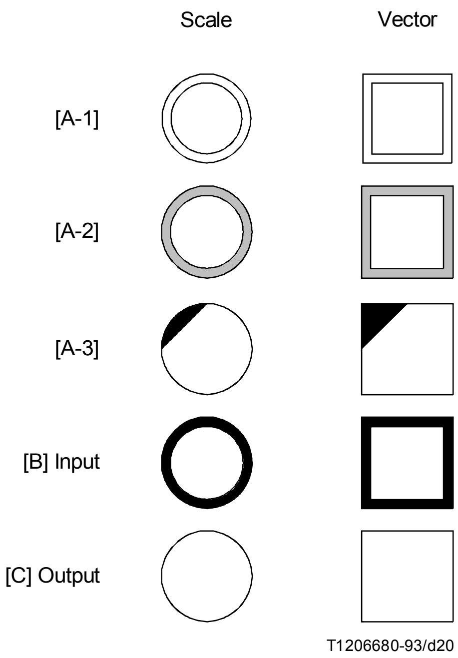


FIGURE 4-2

Graphic symbols used in Figures 4-3 to 4-10

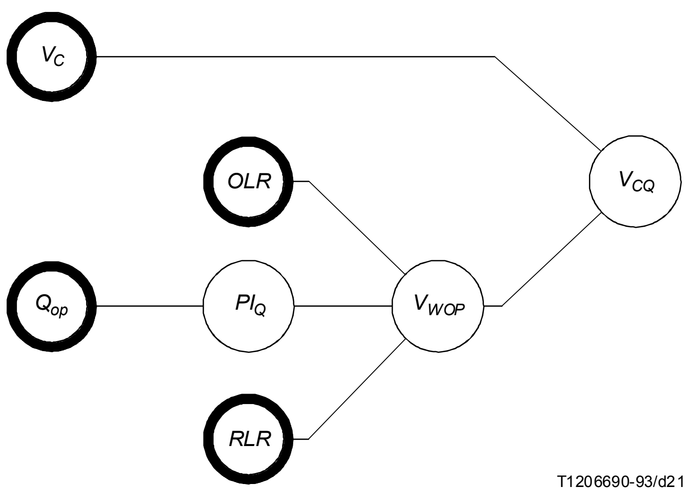


FIGURE 4-3

Block diagram of V_{CQ} calculation

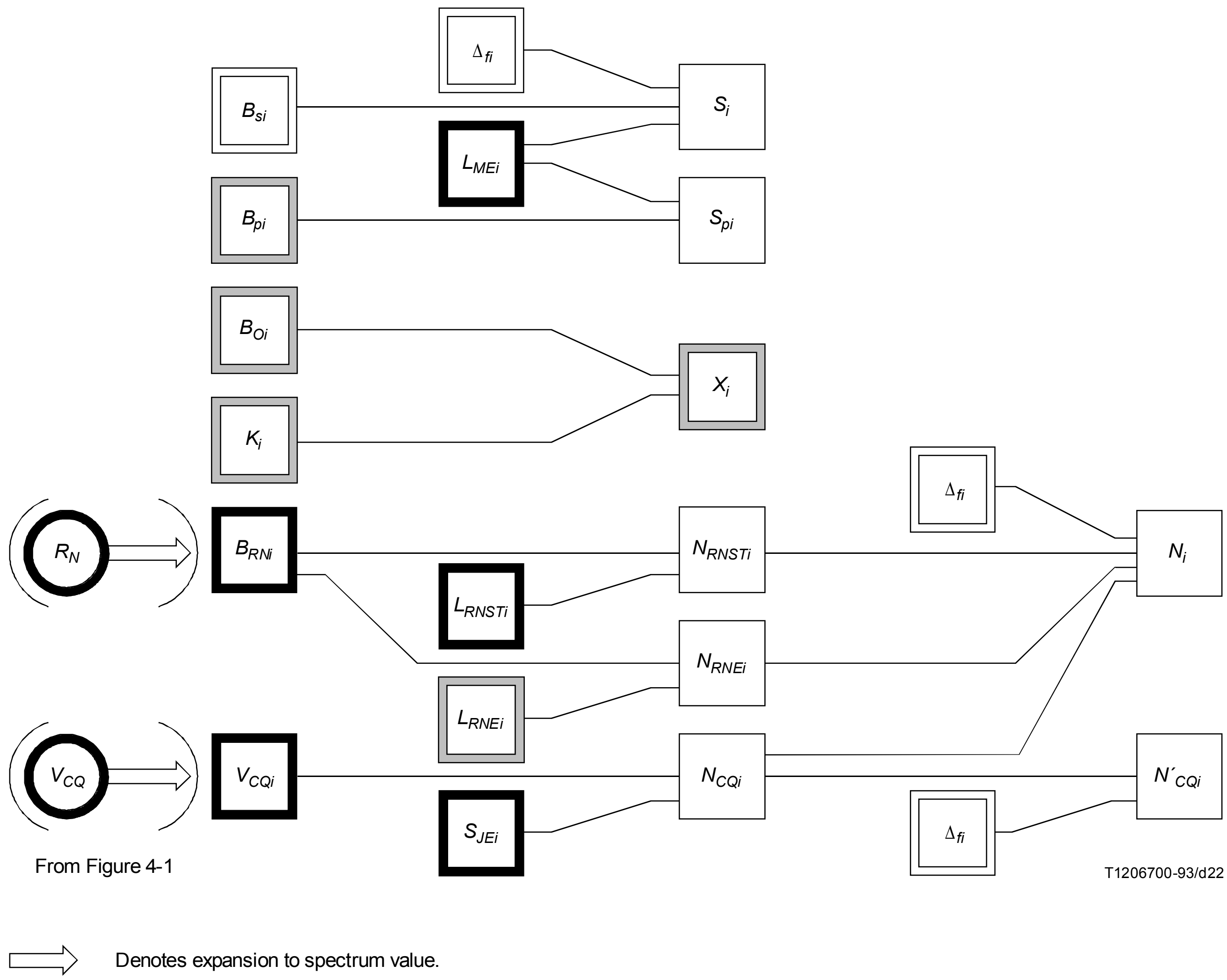


FIGURE 4-4

Block diagram of speech and noise spectrum calculation

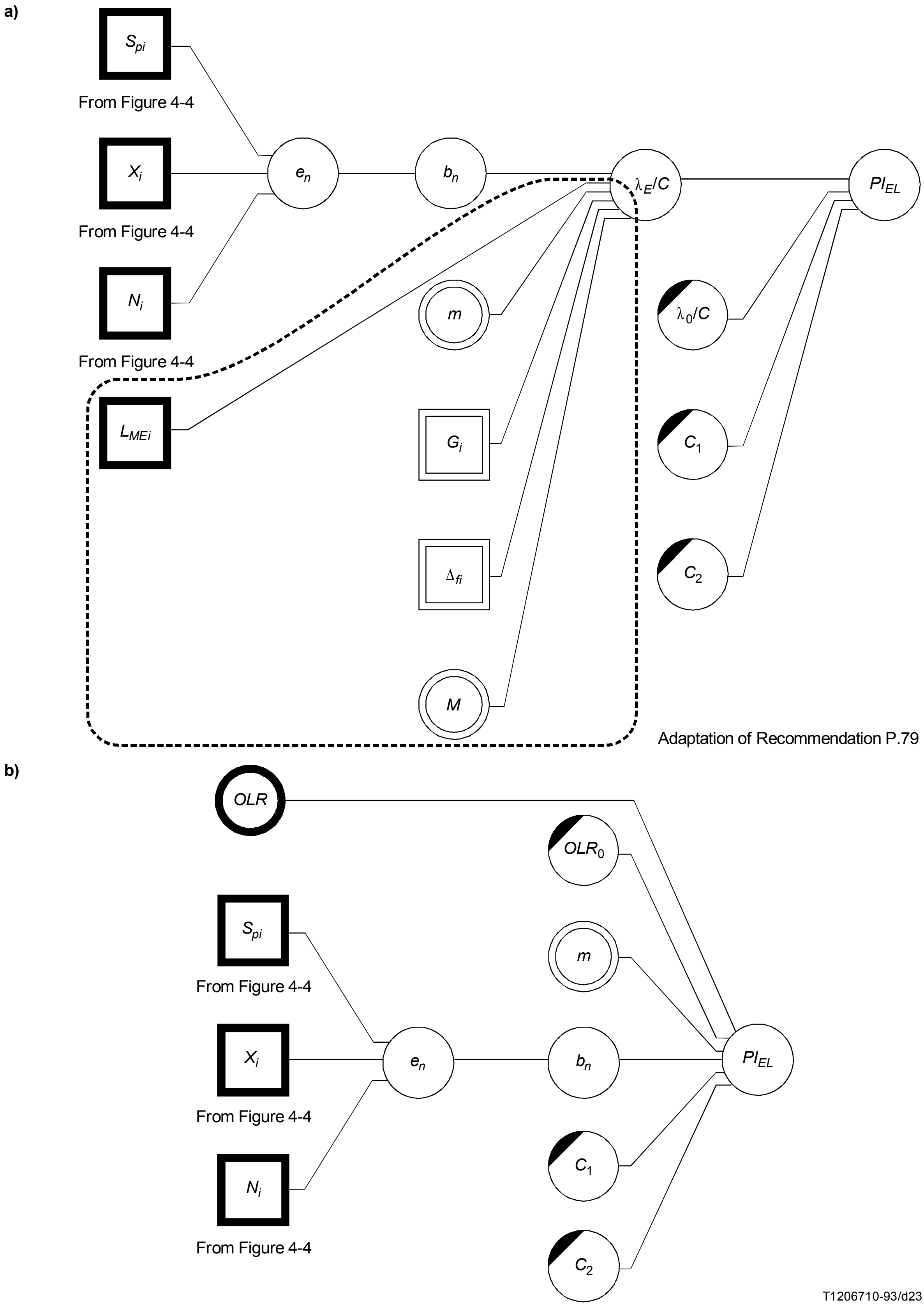


FIGURE 4-5

Block diagram of PI_{EL} calculation

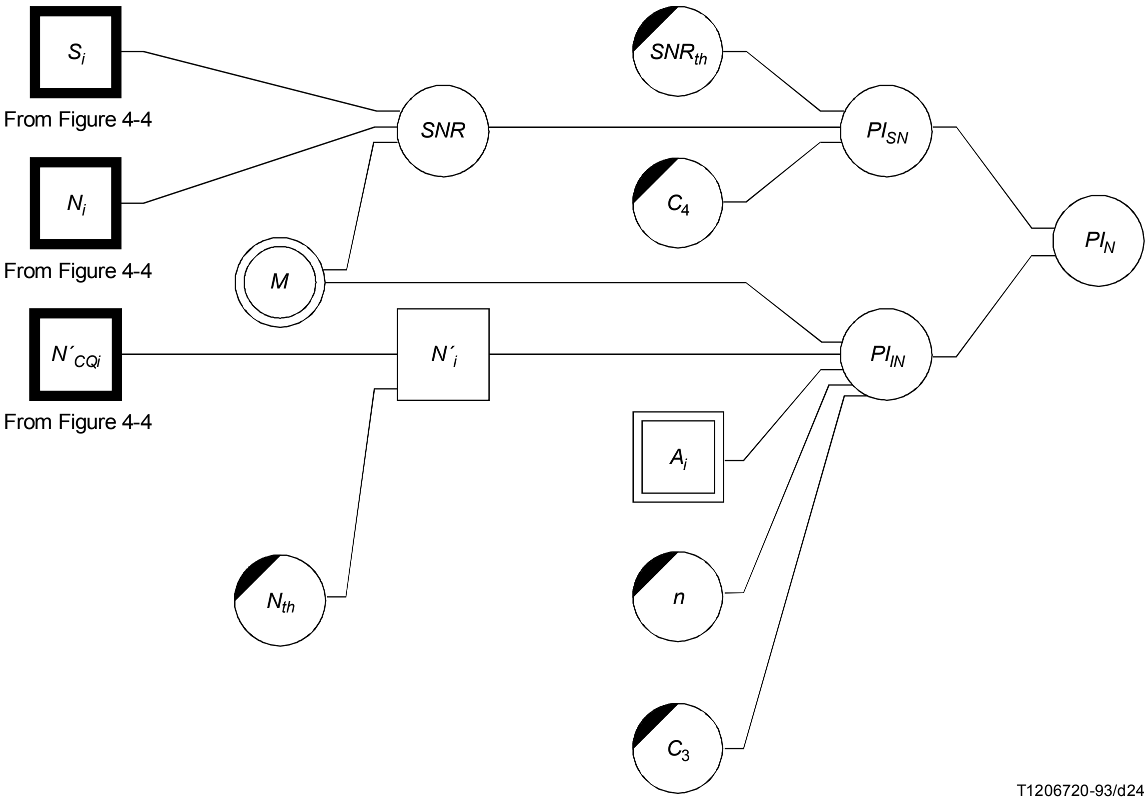


FIGURE 4-6

Block diagram of PI_N calculation

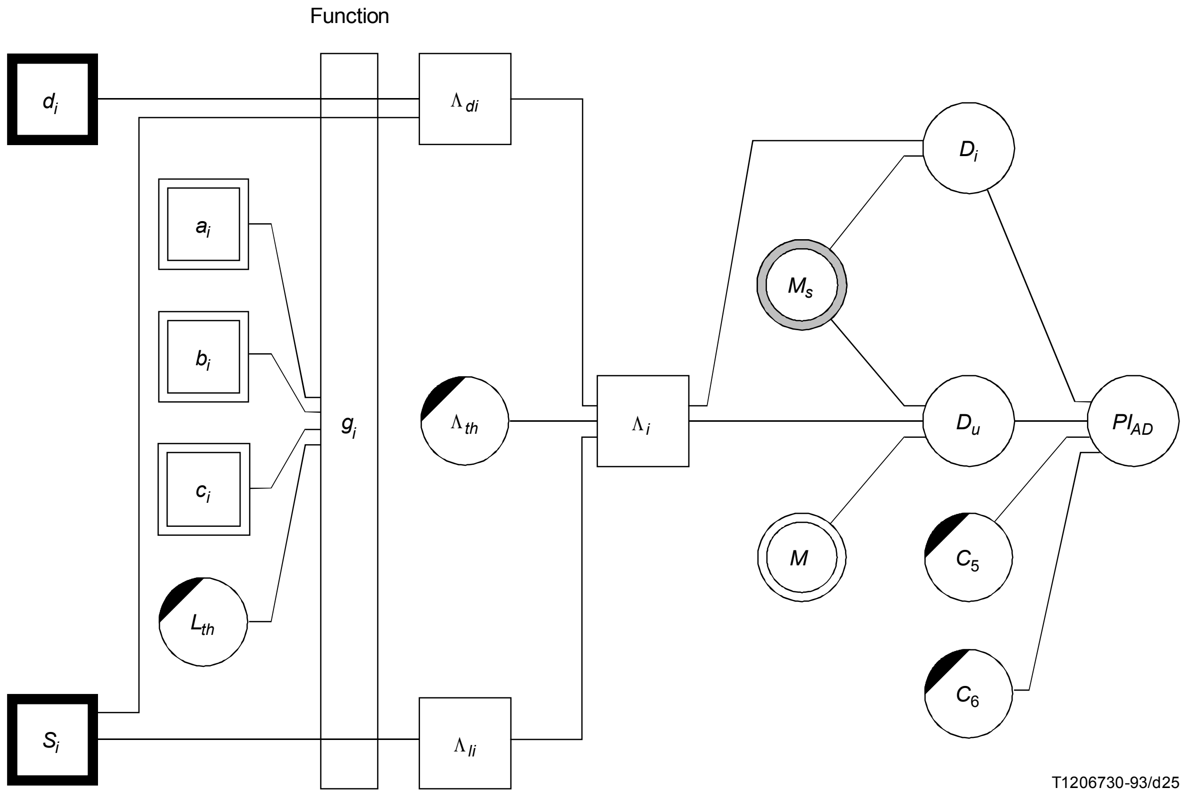


FIGURE 4-7

Block diagram of PI_{AD} calculation

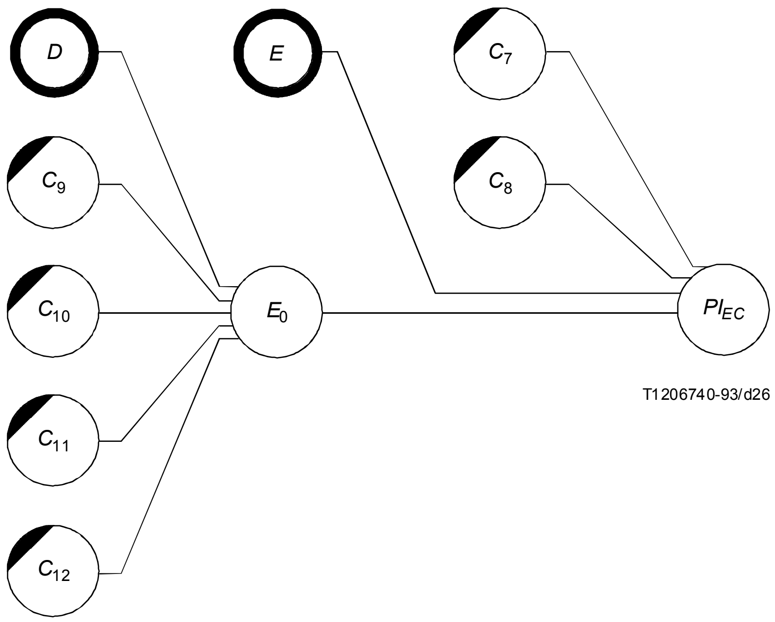


FIGURE 4-8

Block diagram of PI_{EC} calculation

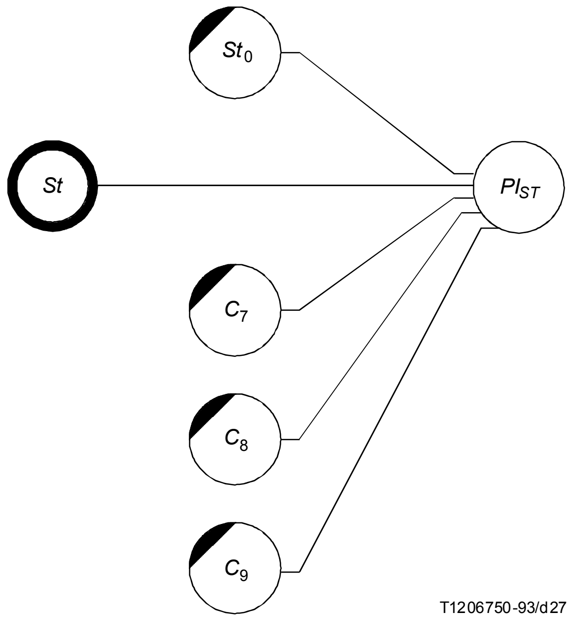


FIGURE 4-9

Block diagram of PI_{ST} calculation

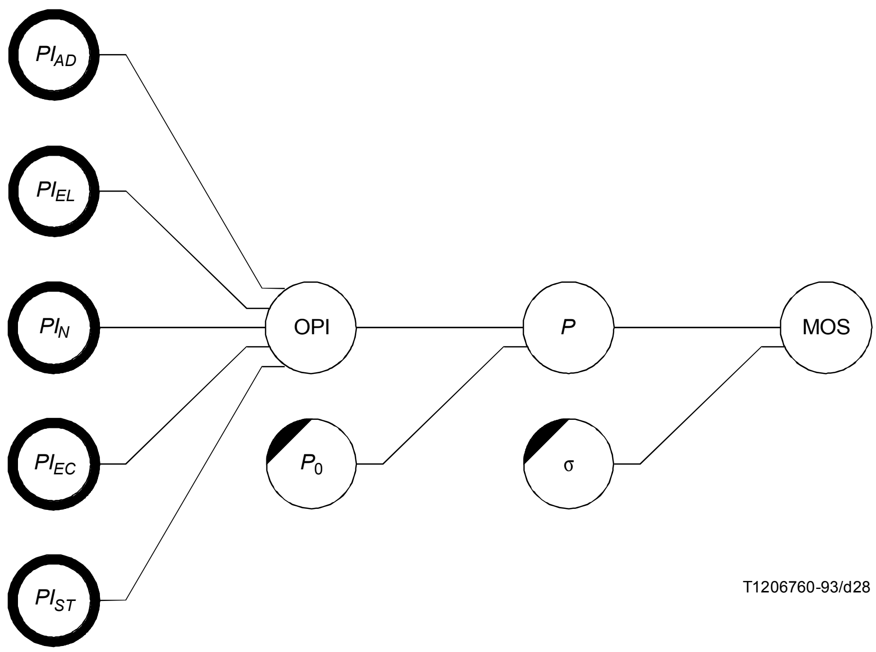


FIGURE 4-10

Block diagram of MOS calculation

TABLE 4-3

Values of a_i, b_i and c_i
(interpolated from [38])

No.	Frequency (Hz)	a_i	b_i	c_i
1	100	–33.5	1.570	– 0.00269
2	125	–25.7	1.500	– 0.00258
3	160	–19.4	1.444	– 0.00248
4	200	–14.7	1.404	– 0.00242
5	250	–10.8	1.362	– 0.00231
6	315	– 7.4	1.314	– 0.00214
7	400	– 4.7	1.259	– 0.00185
8	500	– 3.0	1.205	– 0.00151
9	630	– 1.5	1.141	– 0.00107
10	800	– 0.5	1.064	– 0.00050
11	1000	0.0	1.000	0.00000
12	1250	0.6	0.967	0.00028
13	1600	1.7	0.037	0.00071
14	2000	3.3	0.924	0.00100
15	2500	5.3	0.928	0.00118
16	3150	7.3	0.940	0.00119
17	4000	7.9	0.954	0.00098
18	5000	5.3	0.973	0.00059
19	6300	– 2.6	1.028	0.00013

TABLE 4-4

Model parameters

	No.	Frequency	Δ_t	B_{si}	B_{pi}	X_i	L_{RNE}	$10 \log_{10} G_i$	A_i
Parameter type				[A-1]	[A-2]	[A-2]	[A-2]	[A-1]	[A-1]
Origin				Rec. P.51	$B_{si} + 12$	NTT 1968	NTT 1968	Rec. P.79	ISO
		(Hz)	(Hz)	(dB) 20 µPa/Hz	(dB) 20 µPa/Hz	(dB) 20 µPa/Hz	(dB)	(dB)	(dB)
	1	100	22.4	57.2	69.2	11.0	0.0	–32.63	–19.1
	2	125	29.6	60.0	72.0	8.9	0.0	–29.12	–16.1
	3	160	37.5	62.1	74.1	5.5	0.0	–27.64	–13.4
	4	200	44.7	62.9	74.9	2.2	0.0	–28.46	–10.9
	5	250	57.0	63.0	75.0	0.0	0.0	–28.58	–8.6
	6	315	74.3	62.4	74.4	–3.0	0.7	–31.10	– 6.6
	7	400	92.2	61.0	73.0	–6.0	0.0	–29.78	–4.8
	8	500	114.0	59.3	71.3	–8.0	0.0	–32.68	–3.2
	9	630	149.0	57.0	69.0	–9.5	2.2	–33.21	–1.9
	10	800	184.0	54.2	66.2	–10.3	8.5	–34.14	– 0.8
	11	1000	224.0	51.4	63.4	–11.0	13.5	–35.33	0.0
	12	1250	296.0	48.5	60.5	–11.8	15.5	–37.90	0.6
	13	1600	375.0	45.2	57.2	–13.0	20.0	–38.41	1.0
	14	2000	447.0	42.2	54.2	–16.0	23.7	–41.25	1.2
	15	2500	570.0	39.4	51.4	–19.8	30.0	–41.71	1.3
	16	3150	743.0	36.8	48.8	–23.0	27.0	–45.80	1.2
	17	4000	922.0	34.5	46.5	–26.0	33.5	–43.50	1.0
	18	5000	1140.0	32.7	44.7	–27.0	41.0	–47.13	0.5
	19	6300	1490.0	31.4	43.4	–24.0	50.0	–48.27	– 0.1

NOTE – X_i (= $B_{0i} - k_i$) and L_{RNE} can be input parameters.

TABLE 4-5

Values of estimated constants and coefficients

No.	Related clause	Output	Symbol	Value
1	4.3.2.1 4.3.2.2	PI_{EL}	C_1 C_2 λ_0/c OLR_0	0.0475 0.010 0.780 5.34
2	4.3.2.3	PI_{IN}	N_{th} n C_3	33.0 0.500 0.012
3	4.3.2.3	PI_{SN}	SNR_{th} C_4	7.5 – 0.005
4	4.3.2.4	PI_{AD}	L_{th} C_5 C_6 Λ_{th}	57.5 0.043 0.043 15.0
5	4.3.2.5	PI_{EC}	C_7 C_8 C_9 C_{10} C_{11} C_{12}	13.69 0.01 26.4 2.65 14.00 24.6
6	4.3.2.6	PI_{ST}	C_{13} ST_0	0.00856 9.000
7	4.3.3	MOS	P_0 σ	3.558 0.730

Annex A

(reference to 1.1)

Opinion ratings of transmission impairments

A.1 Introduction

The figures in this annex illustrate the relative effect of typical transmission impairments on opinion ratings. They are based on the transmission rating models described above. The opinion ratings assume a five-category rating scale (excellent, good, fair, poor and bad or unsatisfactory) and the results are presented in terms of the percent of ratings which are good or better (good plus excellent) and poor or worse (poor plus bad). Three equations for the conversion from transmission rating to the opinion ratings are described above in the text of the Supplement. The one which is used in this annex is representative of conversational test results reported to the CCITT by several Administrations during the Study Period 1973-1976.

A.2 Overall loudness rating and circuit noise

Opinion ratings for the combined effects of OLR (L'_e in dB) and circuit noise (N'_c in dBmp) are shown in Figures A.1 and A.2. The circuit noise is referred to a receiving system with an RLR of 0 dB. In these figures the circuit noise equivalent for room noise N'_{Re} is –58.63 dBmp and the bandwidth/slope factor (k_{BW}) is 1; quantization noise, listener echo, talker echo and sidetone are not included.

A.3 Quantization noise from PCM processes

Opinion results for the effect of quantization noise from tandem 7 bit and 8 bit μ-law and A-law PCM processes are shown in Figures A.3 and A.4. These results assume an OLR (L'_e) of 16 dB and a circuit noise (N'_c) of –56 dBmp. Room noise, bandwidth/slope and sidetone assumptions are the same as for A.2. The speech level at the output of a telephone set with a 0 dB SLR is assumed to be –10 VU.

A.4 Bandwidth

The effect on opinion rating as a function of bandwidth between frequencies having 10 dB of loss relative to 1000 Hz is shown in Figures A.5 and A.6. These results assume an OLR (L'_e) of 16 dB, a circuit noise (N'_c) of –56 dBmp, a circuit noise equivalent for room noise (N'_{Re}) of –58.63 dBmp, and lower (S_l) and upper (S_u) slope factors of 2 and 3 respectively. Listener echo, talker echo and sidetone effects are not included.

A.5 Listener echo

The effect of listener echo on opinion ratings is illustrated in Figures A.7 and A.8. In these figures the opinion is plotted (from both the original and alternate models of the supplement) as a function of the weighted listener echo path loss (WEPL) in dB and round-trip listener echo path delay (D_L) in milliseconds. The curves were calculated assuming an OLR (L'_e) of 16 dB, a circuit noise (N'_c) of –56 dBmp, a circuit noise equivalent for room noise (N'_{Re}) of –58.63 dBmp, and a bandwidth/slope factor of 1. Talker echo and sidetone effects are not included.

A.6 Talker echo

Opinion ratings for talker echo are presented in Figures A.9 and A.10 as a function of the OLR of the talker echo path (E) in dB and the round-trip talker echo path delay (D) in milliseconds. Again, the OLR (L'_e) was taken as 16 dB, the circuit noise (N'_c) as –56 dBmp, the circuit noise equivalent of room noise (N'_{Re}) as –58.63 dBmp and the bandwidth/slope factor as 1. Listener echo and sidetone effects are not included.

A.7 Sidetone

Opinion ratings for sidetone are presented in Figures A.11 and A.12 in terms of the sidetone path loss (STMR) in dB and the sidetone response shape in dB/octave. For these curves, impairment levels were selected to provide a constant R_{LN} value typical of toll calls in North America and a range of R_E values which might be encountered on toll calls in North America.

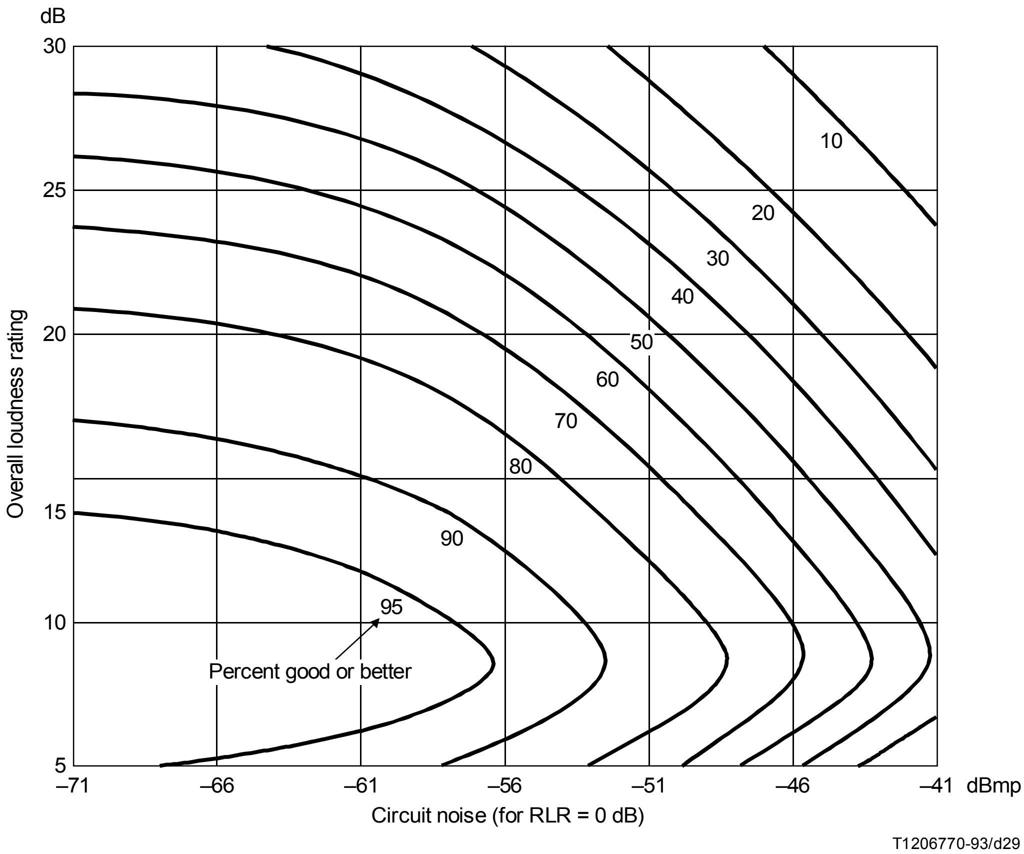


FIGURE A.1

Opinion rating for OLR and circuit noise

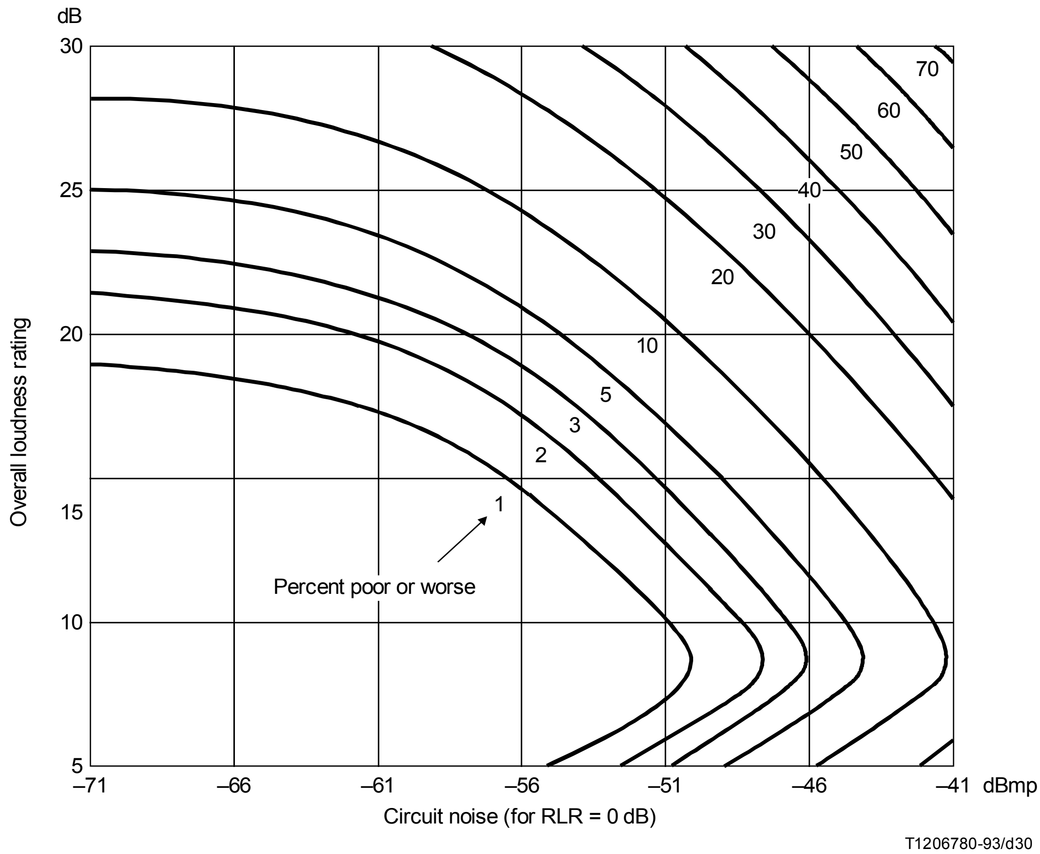


FIGURE A.2

Opinion rating for OLR and circuit noise

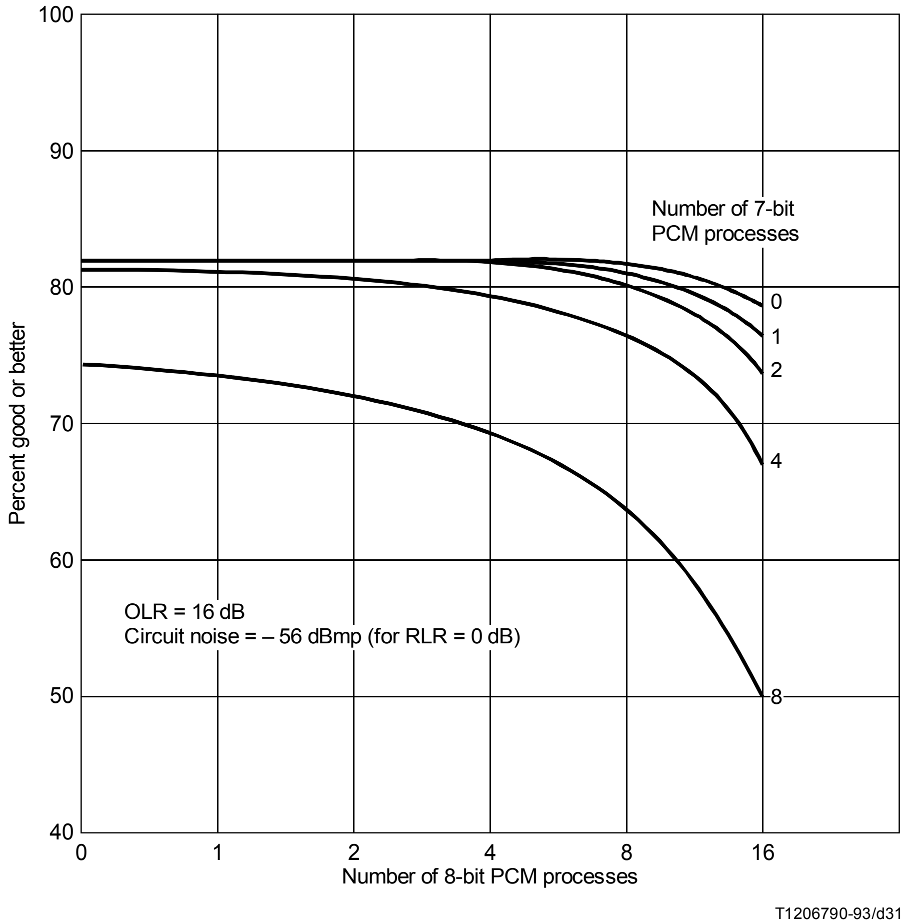


FIGURE A.3

Opinion rating for tandem PCM processes

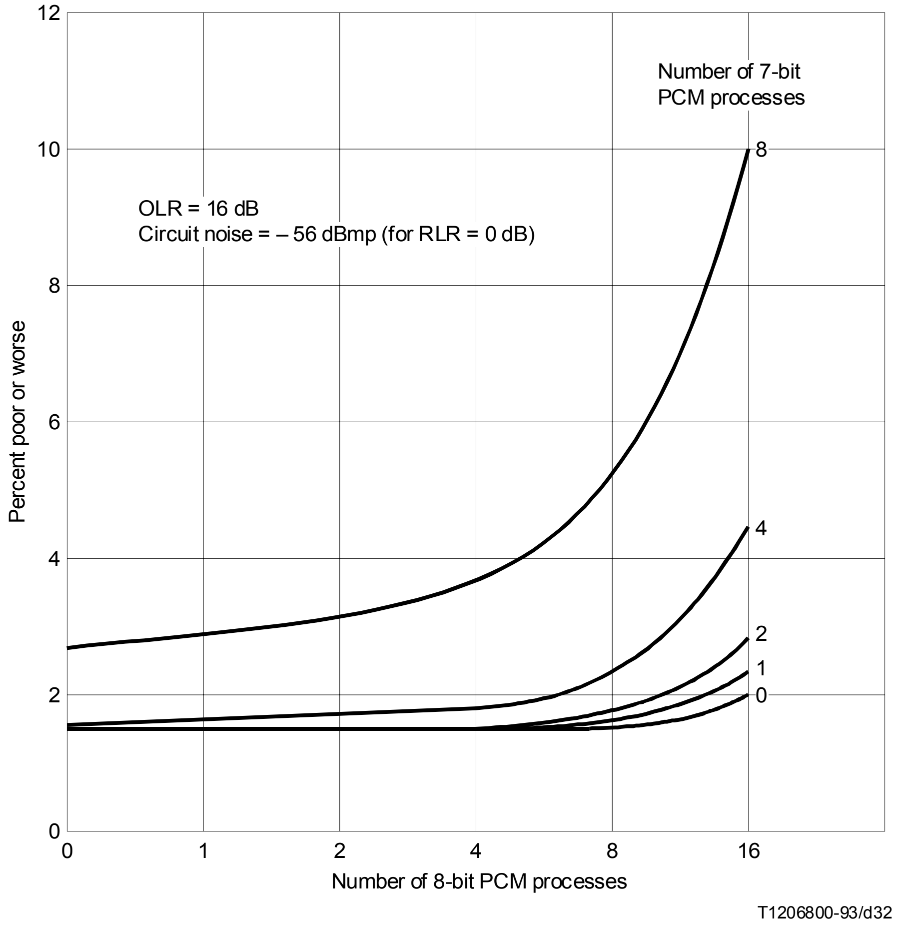


FIGURE A.4

Opinion rating for tandem PCM processes

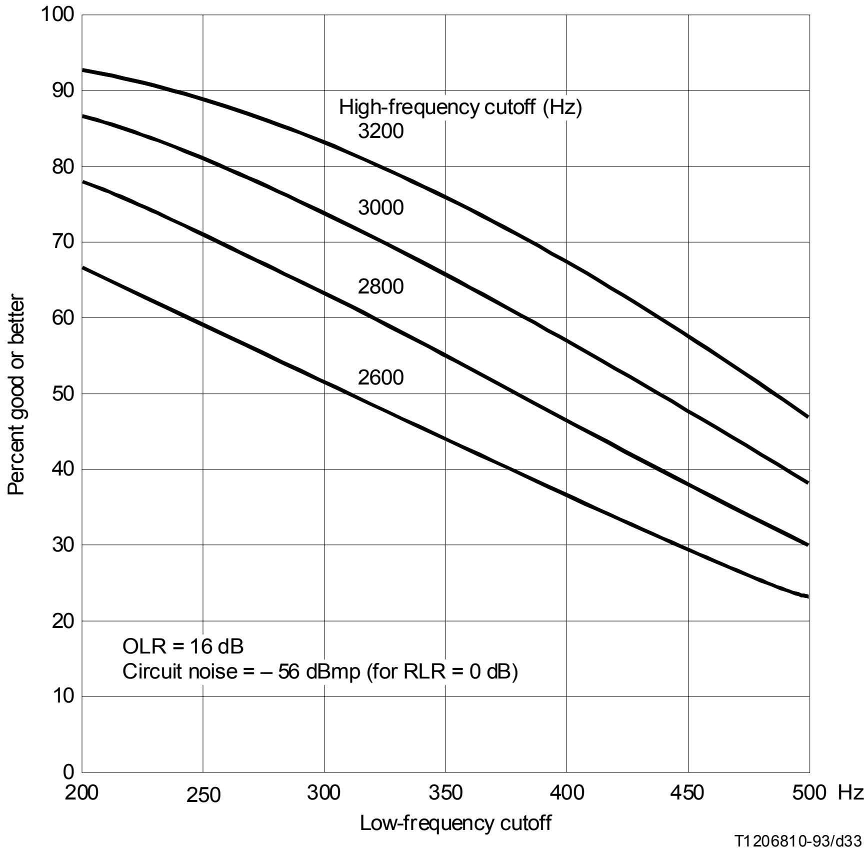


FIGURE A.5

Opinion rating for bandwidth

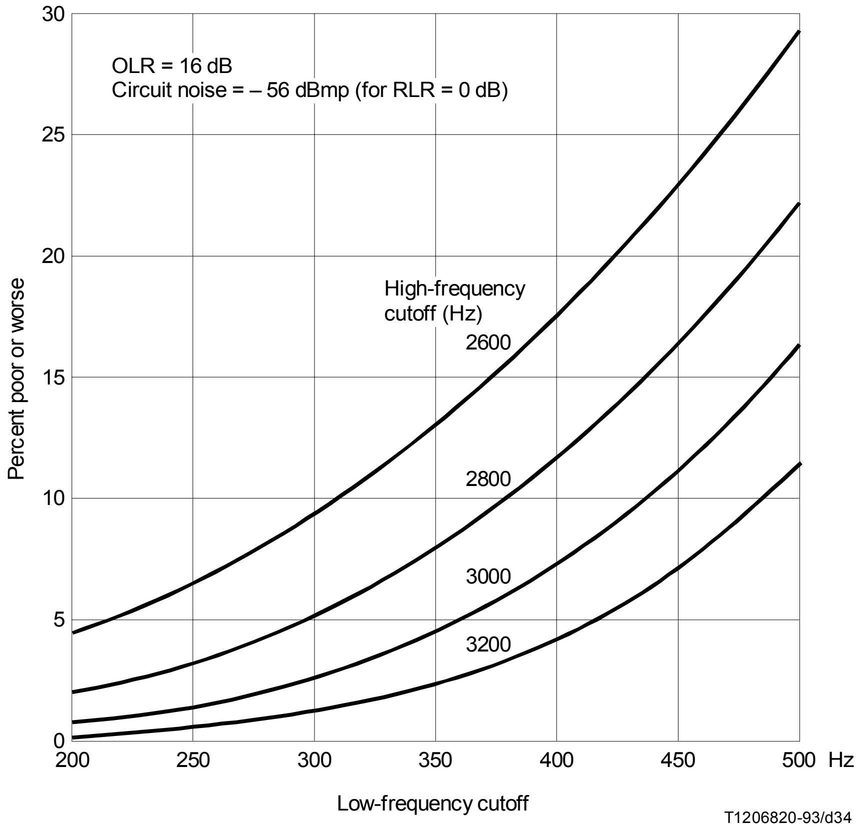


FIGURE A.6

Opinion rating for bandwidth

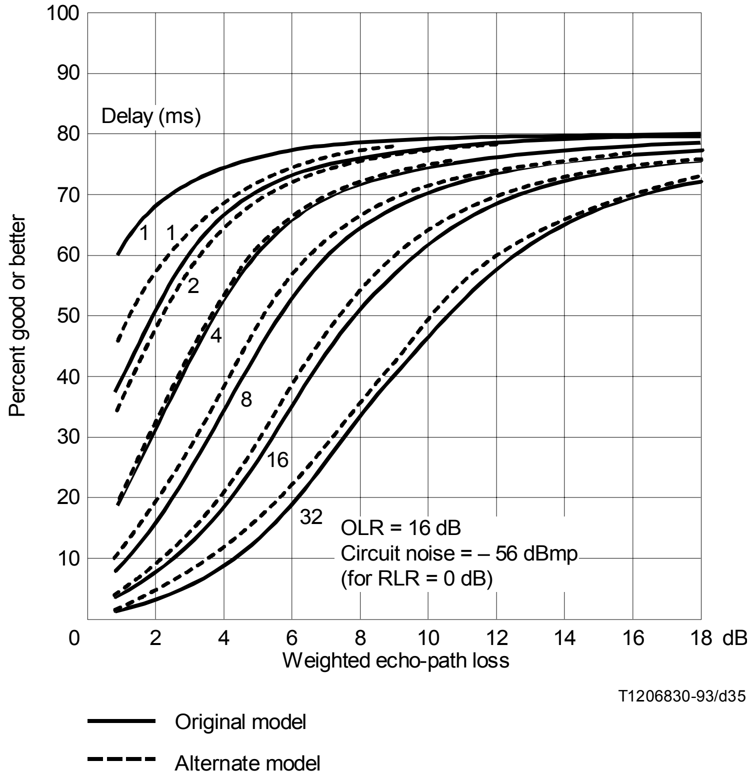


FIGURE A.7

Opinion rating for OLR, circuit noise and listener echo

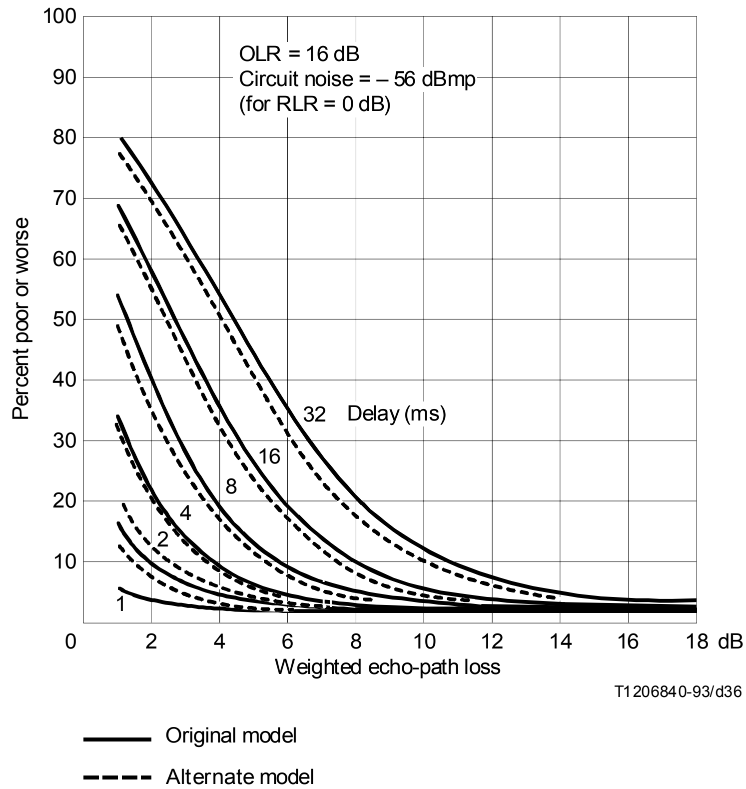


FIGURE A.8

Opinion rating for OLR, circuit noise and listener echo

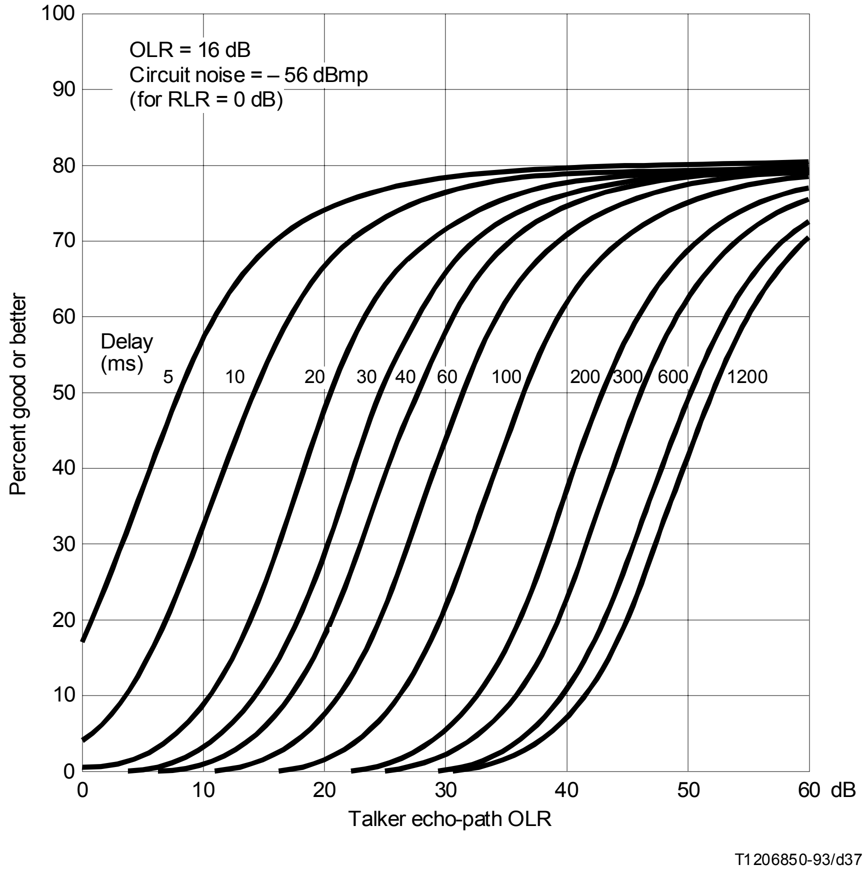


FIGURE A.9

Opinion rating for OLR, circuit noise and talker echo

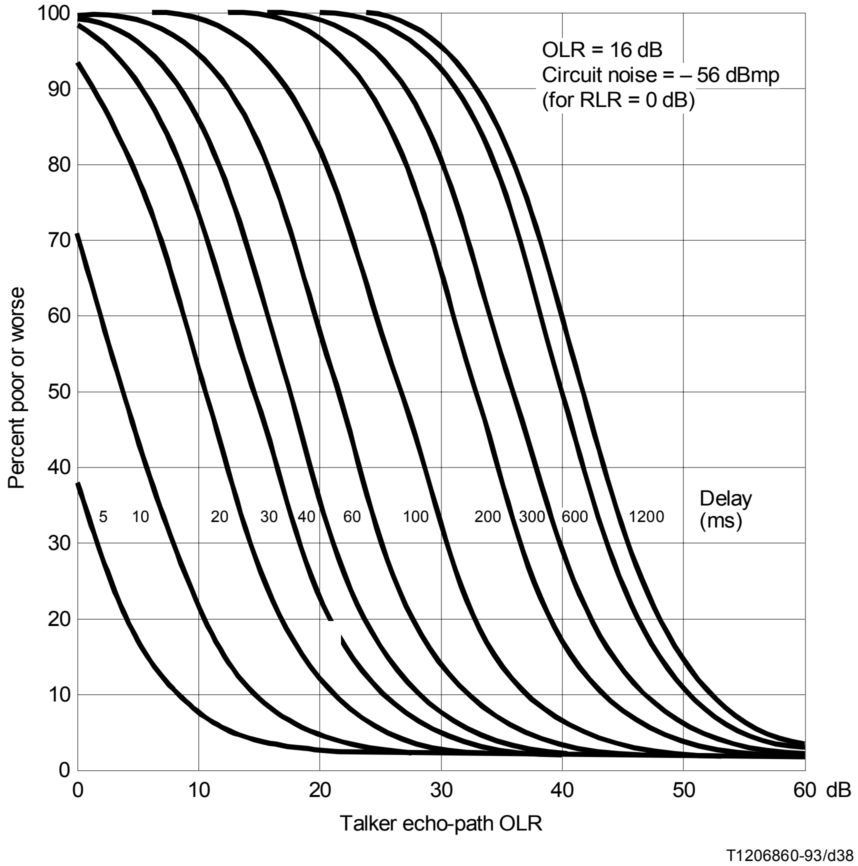


FIGURE A.10

Opinion rating for OLR, circuit noise and talker echo

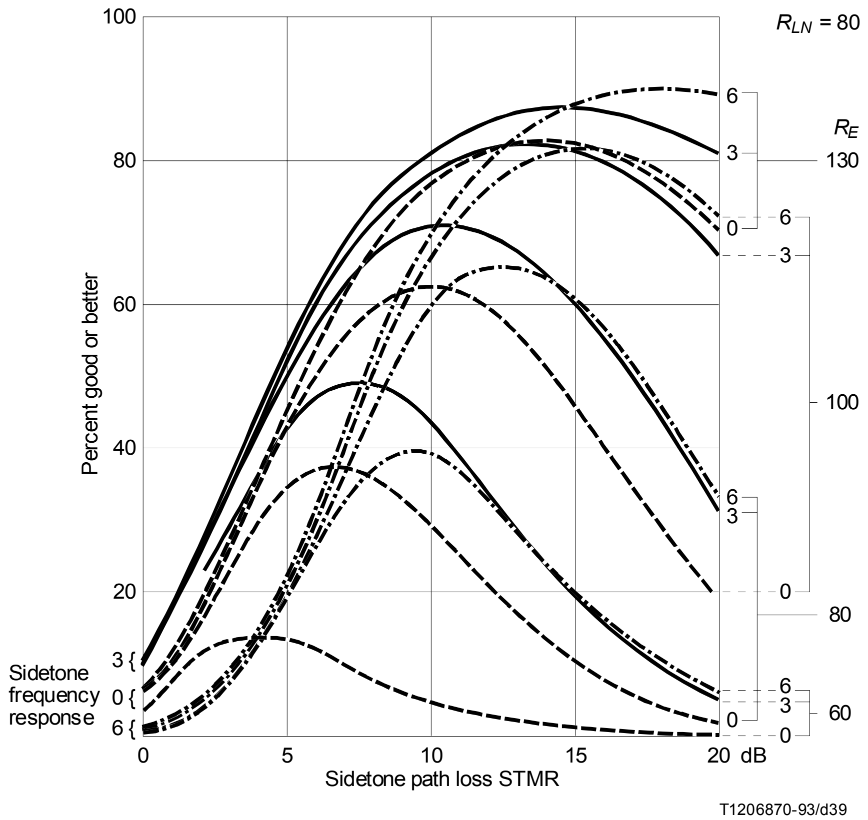


FIGURE A.11

Opinion rating for OLR, circuit noise, talker echo and sidetone

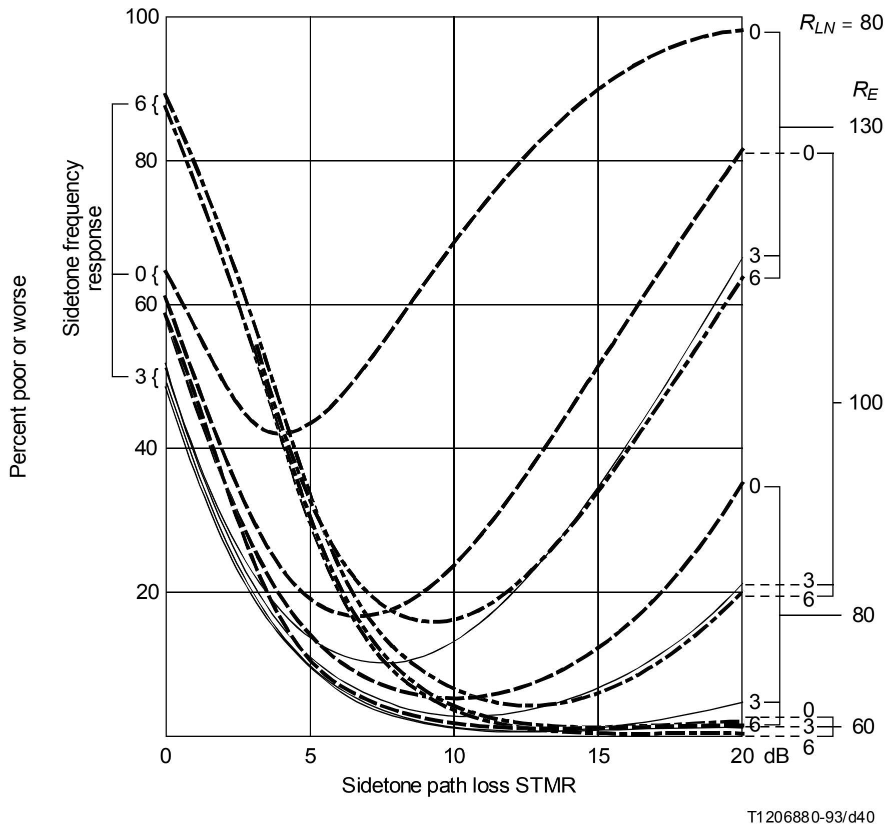


FIGURE A.12

Opinion rating for OLR, circuit noise, talker echo and sidetone

Annex B

(reference to 2.9)

Calculated transmission performance of telephone networks

B.1 Introduction

This annex is intended to give examples of results from the subjective model which is incorporated in the BT CATNAP (Computer-Aided Telephone Network Assessment Program) program. CATNAP comprises this model and a transmission calculation section which enables elements of a connection to be entered as readily identifiable items, e.g. lengths of cable, feed bridges etc. These results are examples of calculations for various "hypothetical reference connections" (HRCs) which might arise in the network or would be of use to planners.

The loudness ratings quoted are calculated according to Recommendation P.79, using the frequency bands from 200 Hz to 4 kHz. The opinion scores, Y_{LE} and Y_C, are on a scale of 0 to 4, representing the listening effort and conversation opinion scales (see Recommendation P.80). The values of line current shown with the results are determined by the program which decides from the characteristics of the local telephone system which of a number of standard line currents is appropriate, and hence which values of the telephone instrument characteristics should be used. The program also gives speech levels for controlled talking conditions (V_L) and under conversational conditions (V_C). These and the loudness ratings are referred to the interfaces (NI and FI) shown in the figures below.

These results are for the model as it stands at present (1983 version). Research is continuing to improve the correlation of calculated and experimental results, so the model is liable to modification.

B.2 HRC 1 – Own exchange call (see Figure B.1)

This is a symmetrical connection, with average length customers' lines. The sidetone suppression is fairly good, and room noise and circuit noise levels are low. The conversation opinion score is good, but the small overall loss means that the connection is louder than preferred. A slightly quieter connection would give a better opinion score.

B.3 HRC 2 – Limiting national call (see Figure B.2)

These two HRCs are both symmetrical and comprise BT limiting local lines of 1000Ω/10 dB, 4.5 dB local junctions and two 4-wire junctions each with 3.5 dB loss, which are the limits set by the BT transmission plan (given in [22]).

HRC 2 (a) uses 0.5 mm copper local lines, which provide much better sidetone matching than the 0.9 mm copper lines of HRC 2 (b). The change in sidetone level (> 10 dB) causes a drop in the conversation opinion score from 1.9 to 0.8 (from fair to poor).

B.4 HRC 3 – Long distance call with a PCM junction (see Figure B.3)

The overall loss of this connection (OLR = 13.4 dB) is much less than for HRC 2. The local lines are average length of 0.5 mm copper which give reasonably good sidetone matching, and there is now only one local junction. This is a 4-wire 3 dB PCM junction. This is entered as a single item, characterised by the terminating and balance impedances of the 2/4-wire terminating sets, the matched loss in each direction and the phase delay round the loop. Quantizing noise is negligible for the input speech levels calculated by CATNAP for this connection.

The connection is symmetrical in transmission loss but a small difference in the sidetone level has given slightly different conversation opinion scores at the two ends.

B.5 HRC 4 – Asymmetry of transmission loss (see Figure B.4)

A number of calculations have been done for this HRC to show the effect of varying the degree of asymmetry. The curves shown are not fitted curves, but simply join the marked points on the graph. They show the effect on the conversation opinion score and conversational speech voltage of varying the transmission loss in one direction only (from near end to far end). The loss from far to near is kept constant, so the opinion of the near end customer is much less affected. It is suspected that the speech voltage curves are too divergent and further research is needed in this area, but the opinion curves show similar trends to the results produced by Boeryd [30].

The sidetone level was virtually unaffected by the change in transmission loss.

B.6 HRC 5 – Effect of room noise (see Figure B.5)

The calculations done for this HRC demonstrate the effect of changing the level of room noise for a customer with a loud sidetone path (near end) and one with a quiet sidetone path (far end). As for HRC 4, the computed points are simply joined to form the line.

B.7 HRC 6 – Effect of circuit noise and bandlimiting (see Figure B.6)

This is a connection using 4-wire reference telephones, enabling sidetone to be controlled. The STMR is kept at 20 dB, at which level most customers would not detect it.

Such a connection can be used to investigate the effects of particular transmission impairments varied independently. Here it has been used to demonstrate the effect on the listening effort and conversation opinion scores of the level of injected circuit noise and band limiting (lowpass) over a range of losses likely to occur in telephone networks.

As for the previous curves the computed points are simply joined to form a line.

B.8 HRC 7 – Multiple calculations with random selection of items (see Figure B.7)

CATNAP is intended to help assess telephone network proposals rather than single connections. The program can perform multiple calculations on a group of connections or on a single connection with random selection of elements from a database.

Here random selection is made of the customers' lines out of a database derived from a survey of 1800 existing lines. This enables the performance of a particular element to be tested for a range of conditions which would arise in the actual network. Since the survey reflects the distribution of lengths and gauges in the actual network, this method of assessment gives a more accurate picture of the performance in the existing network.

For this example only a few calculations have been done to demonstrate the facility and so the results have been printed. This is not practical for large numbers of calculations, when the results are stored and can be processed as desired, e.g. by plotting the distribution or by statistical analysis.

The line number and radial distance have been given for both ends of each calculation.

B.9 HRC 8 – Example of the use of CATNAP to meet a design criterion (see Figure B.8)

This is intended to give an example of the use of CATNAP in the design of individual network components to meet design targets.

With the introduction of electronic telephones the designer has a freer choice of values for the telephone instrument characteristics, e.g. the value of the line impedance which must be connected to the telephone instrument to give full sidetone suppression (Z_{SO}).

An iterative procedure can lead to preferred values for Z_{SO}. As examples, calculations have been done for a standard BT 706 and a 706 with some trial values for Z_{SO} on BT limiting lengths of local copper cable of standard gauges, and an average length of 0.5 mm cable. For one of the trial sets of values which looks possible from these results and for a standard 706 instrument, a set of 40 calculations were done with a random selection of local lines from the database of 1800 used for HRC 7. These results are given in terms of the mean and standard deviation of the distribution of STMRs. From this it can be seen that the trial values do give a better performance on average, although the performance is worse on 0.63 mm and 0.9 mm limiting lines, since these are less common in the local network than 0.5 mm.

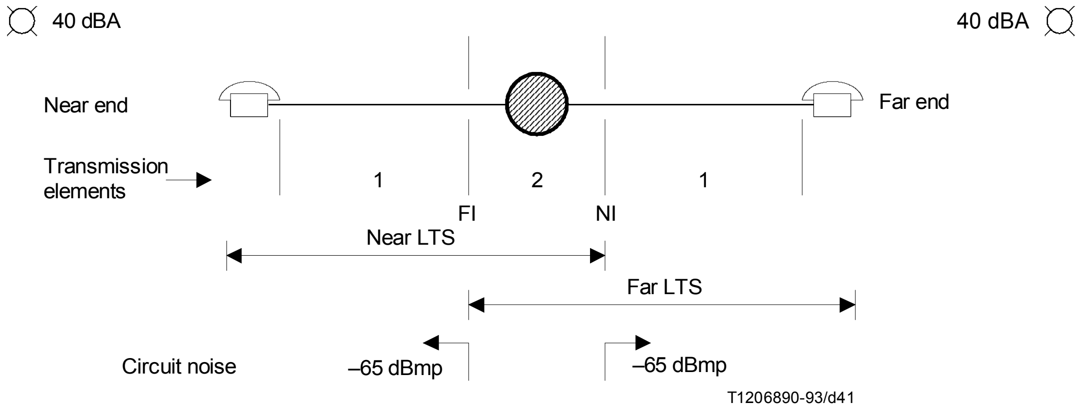
As a design tool, the program could be used further to verify the improvement in performance, to check the effects of tolerances and to consider possible improvements to these values.

B.10 HRC 9 – Effect of varying line length (see Figure B.9)

This HRC is identical to HRC 2 except for the gauge of cable. In this case 0.63 mm copper cable is used. Its length is varied from zero to 10 km, which is beyond the BT limiting length (7.2 km).

The results are shown as curves of conversation opinion score, OLR and conversational speech voltage against line length. As before, the computed points are simply joined to form a line.

The calculations on this HRC have been included to demonstrate the "inverse" use of CATNAP. The limits on OLR are known (from the transmission plan) and so these runs could be used to show what range of cable lengths are acceptable. The facility for calculating the performance in terms of conversation opinion score makes it possible to specify performance limits in terms of this, which is closer to the real performance than limits set in terms of loudness ratings.



Transmission elements

Telephone instruments are BT Type No. 706

1 Unloaded cable 1.6 km of 0.5 mm (168 ohms/km, 50 nF/km)

2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)

Near end			*IL*	=	64
STMR	=	9.02	SLR	=	4.76
RLR	=	–5.15	OLR	=	–0.27
Y_{LE}	=	3.48	V_L	=	–18.24
Y_C	=	3.15	V_C	=	–22.69
RN	=	40.00	ICN	=	–65.00

IL	=	64	*Far end*		
OLR	=	–0.27	RLR	=	–5.15
SLR	=	4.76	STMR	=	9.02
V_L	=	–18.24	Y_{LE}	=	3.48
V_C	=	–22.69	Y_C	=	3.15
ICN	=	–65.00	RN	=	40.00

IL	Line current (mA)
SLR	Sending loudness rating (dB)
RLR	Receiving loudness rating (dB)
OLR	Overall loudness rating (dB)
STMR	Masked sidetone loudness rating (dB)
Y_{LE}	Listening effort score
Y_C	Conversation opinion score
V_L	Speech voltage at interface (dBV) under controlled talking conditions
V_C	Speech voltage at interface (dBV) under conversational conditions
RN	Level of room (environmental) noise (dBA), Hoth spectrum
ICN	Level of injected circuit noise referred to a 0 dB RLR receiving end
NI	Near interface
FI	Far interface
LTS	Local telephone system

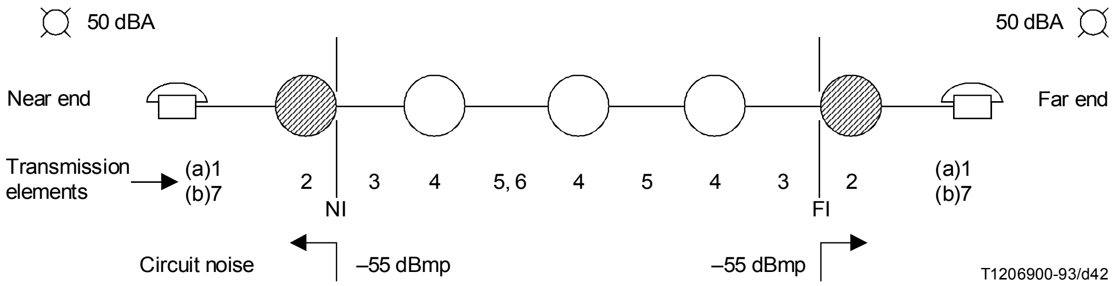
NOTES

1 Room noise has Hoth spectrum.

2 The OLR printed in the left column is for near to far, the OLR in the right column is for far to near.

FIGURE B.1

HRC 1 – Own exchange call



Transmission elements

Telephone instruments are BT Type No. 706

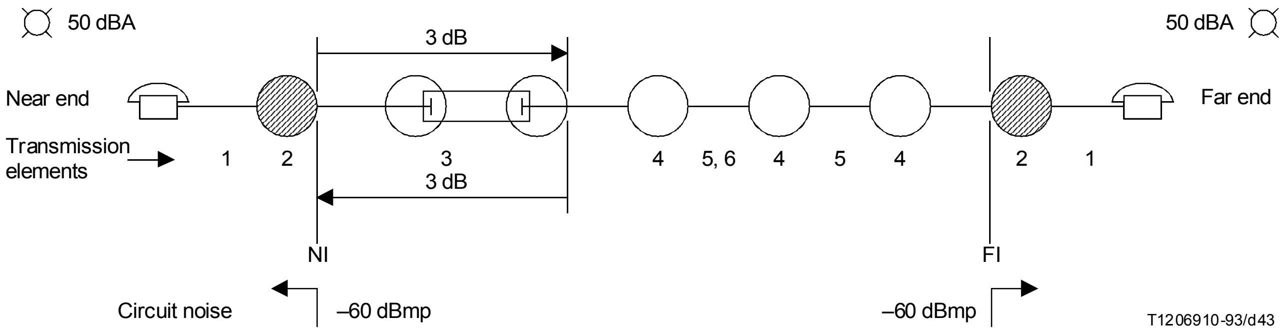
1 Unloaded cable 6 km of 0.5 mm (168 ohms/km, 50 nF/km)
2 Stone feed bridge (2 × 200 Ω, 2 + 2 µF, 50 V)
3 Loaded junction 19.6 km of 0.9 mm, 88 mH, @ 1.83 km
4 Transformer feed bridge (50 V)
5 Attenuation 3.5 dB, frequency independent, 600 Ω
6 Channel filtering 300 Hz-3.4 kHz, 600 Ω
7 Unloaded cable 10 km of 0.9 mm (55 ohms/km, 50 nF/km)

Near end			*IL*	=	32	*IL*	=	32	*Far end*		
STMR	=	11.19	SLR	=	8.21	OLR	=	25.07	RLR	=	–1.32
RLR	=	–1.32	OLR	=	25.07	SLR	=	8.21	STMR	=	11.19
Y_{LE}	=	1.98	V_L	=	–21.40	V_L	=	–21.40	Y_{LE}	=	1.98
Y_C	=	1.86	V_C	=	–22.46	V_C	=	–22.46	Y_C	=	1.86
RN	=	50.00	ICN	=	–55.00	ICN	=	–55.00	RN	=	50.00

Near end			*IL*	=	50	*IL*	=	50	*Far end*		
STMR	=	–0.14	SLR	=	6.62	OLR	=	24.04	RLR	=	–2.05
RLR	=	–2.05	OLR	=	24.04	SLR	=	6.62	STMR	=	–0.14
Y_{LE}	=	1.72	V_L	=	–19.75	V_L	=	–19.75	Y_{LE}	=	1.72
Y_C	=	0.81	V_C	=	–21.52	V_C	=	–21.52	Y_C	=	0.81
RN	=	50.00	ICN	=	–55.00	ICN	=	–55.00	RN	=	50.00

FIGURE B.2

HRC 2 – Limiting national call



Transmission elements

Telephone instruments are BT Type No. 706

1 Unloaded cable 1.6 km of 0.5 mm (168 ohms/km, 50 nF/km)

2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)

3 PCM system 3 dB up to 3.4 kHz, 600 Ω

4 Transformer feed bridge (50 V)

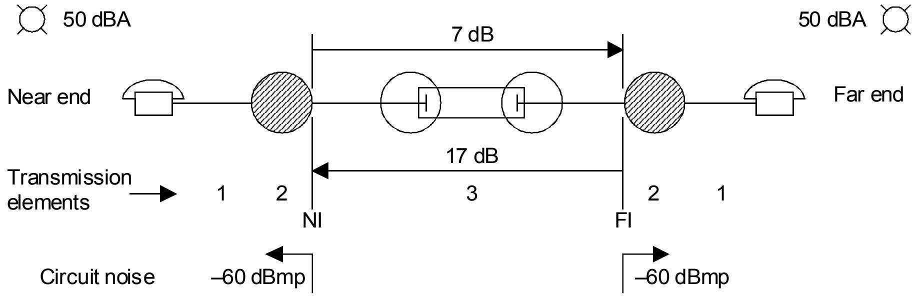
5 Attenuation 3.5 dB, frequency independent, 600 Ω

6 Channel filtering 300 Hz-3.4 kHz, 600 Ω

Near end			*IL*	=	64	*IL*	=	64	*Far end*		
STMR	=	9.31	SLR	=	4.95	OLR	=	13.38	RLR	=	–4.95
RLR	=	–4.95	OLR	=	13.38	SLR	=	4.95	STMR	=	8.55
Y_{LE}	=	3.34	V_L	=	–18.60	V_L	=	–18.60	Y_{LE}	=	3.34
Y_C	=	2.73	V_C	=	–22.19	V_C	=	–22.45	Y_C	=	2.75
RN	=	50.00	ICN	=	–60.00	ICN	=	–60.00	RN	=	50.00

FIGURE B.3

HRC 3 – Long distance call with a PCM junction

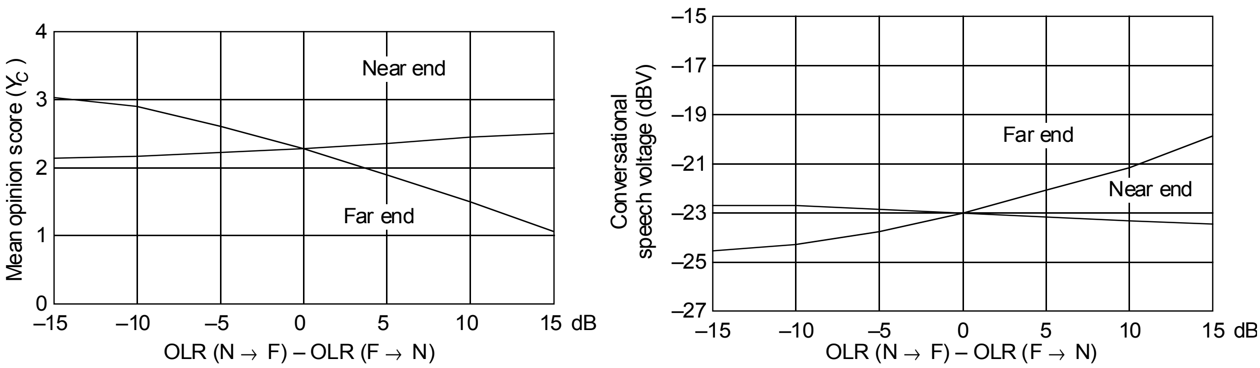


Transmission elements

Telephone instruments are BT Type No. 706

1 Unloaded cable 6 km of 0.5 mm (168 ohms/km, 50 nF/km)
2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
3 FDM system loss as shown up to 3.4 kHz, 600 Ω

Near end			*IL*	=	32	*IL*	=	32	*Far end*		
STMR	=	13.89	SLR	=	8.21	OLR	=	14.35	RLR	=	–1.31
RLR	=	–1.31	OLR	=	24.22	SLR	=	8.21	STMR	=	13.89
Y_{LE}	=	2.62	V_L	=	–21.40	V_L	=	–21.40	Y_{LE}	=	3.40
Y_C	=	2.18	V_C	=	–22.69	V_C	=	–24.22	Y_C	=	2.90
RN	=	50.00	ICN	=	–60.00	ICN	=	–60.00	RN	=	50.00

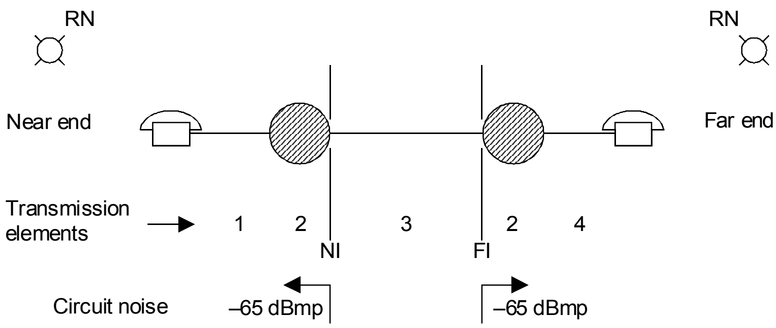


T1206930-93/d44

NOTE – The results shown in the curves are for the same connection with the loss from near to far in the FDM system varied from 2 dB to 32 dB. The loss from far to near was kept at 17 dB.

FIGURE B.4

HRC 4 – Effect of asymmetry of transmission loss



Transmission elements

Telephone instruments are BT Type No. 706

1 Unloaded cable 10 km of 0.9 mm (55 ohms/km, 50 nF/km)
2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
3 Attenuation 20 dB, frequency independent, 600 Ω
4 Unloaded cable 6 km of 0.5 mm (168 ohms/km, 50 nF/km)

Near end			*IL*	=	50	*IL*	=	32	*Far end*		
STMR	=	0.28	SLR	=	6.62	OLR	=	24.69	RLR	=	–1.32
RLR	=	–2.04	OLR	=	25.53	SLR	=	8.82	STMR	=	13.66
Y_{LE}	=	2.63	V_L	=	–19.75	V_L	=	–21.40	Y_{LE}	=	2.87
Y_C	=	2.17	V_C	=	–24.83	V_C	=	–22.67	Y_C	=	2.17
RN	=	40.00	ICN	=	–65.00	ICN	=	–65.00	RN	=	40.00

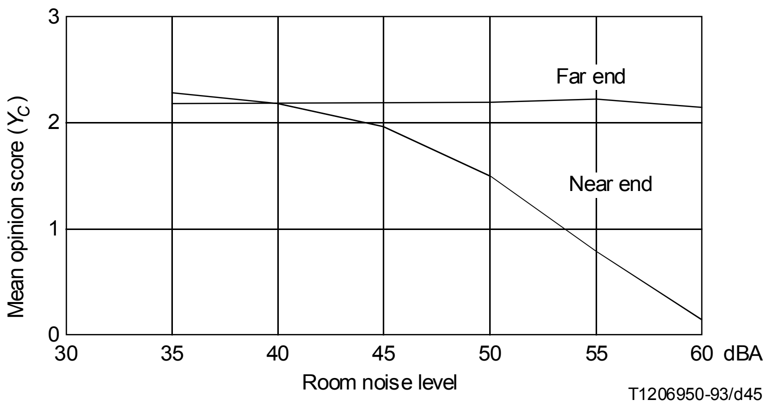
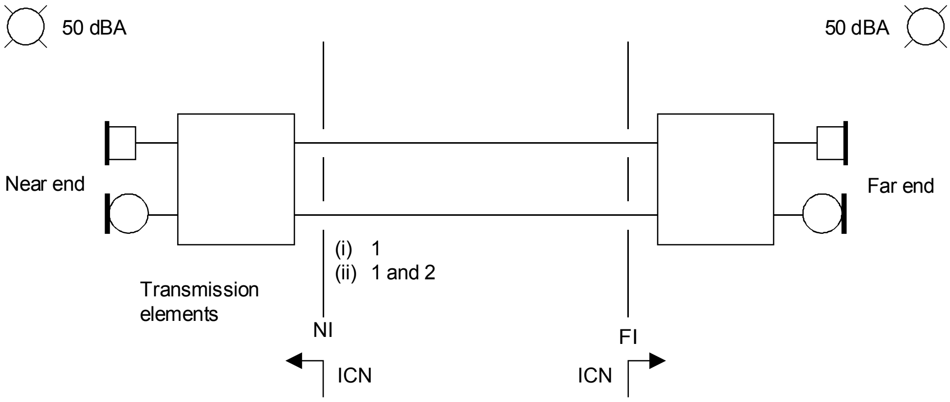


FIGURE B.5

HRC 5 – Effect of room noise level

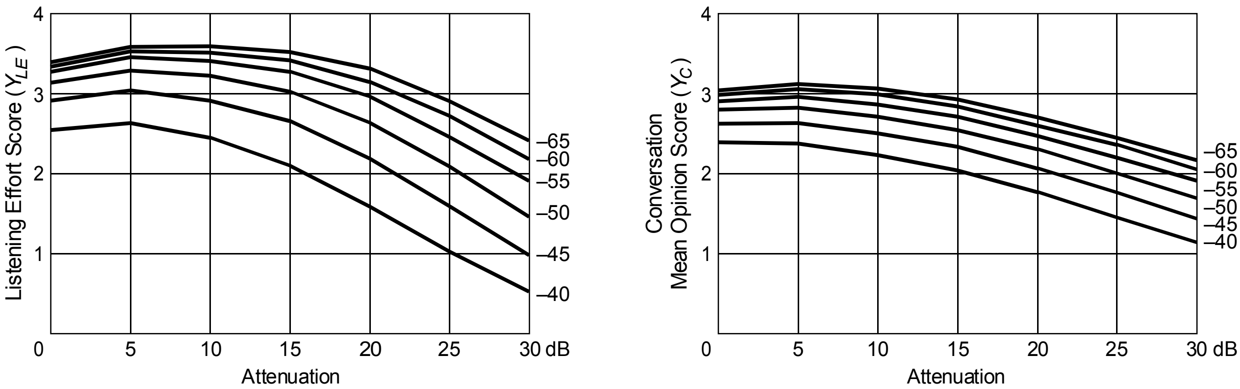


Transmission elements

Telephone instruments are Intermediate Reference Systems (see Recommendation P.48) with a sidetone path of STMR = 20 dB.

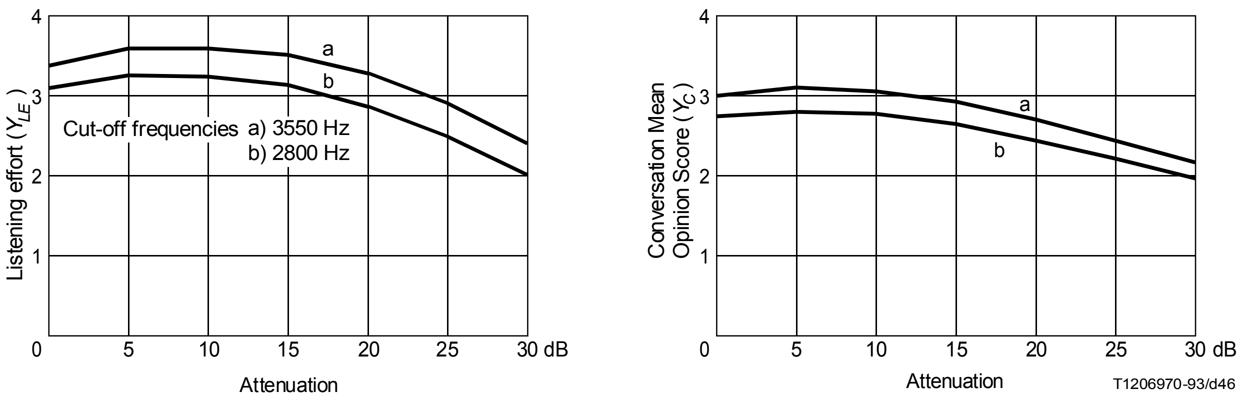
1 Attenuation 0-30 dB, frequency independent, 600 Ω

2 Filtering 600 ohms, (a) 0-3.55 kHz
(b) 0-2.80 kHz



NOTE – These curves show the effect on Y_{LE} and Y_C of changing the level of injected circuit noise from –65 dBmp to –40 dBmp, referred to a 0 dB RLR.

a) Effect of injected circuit noise level and overall loss on the listening effort and conversation opinion scores

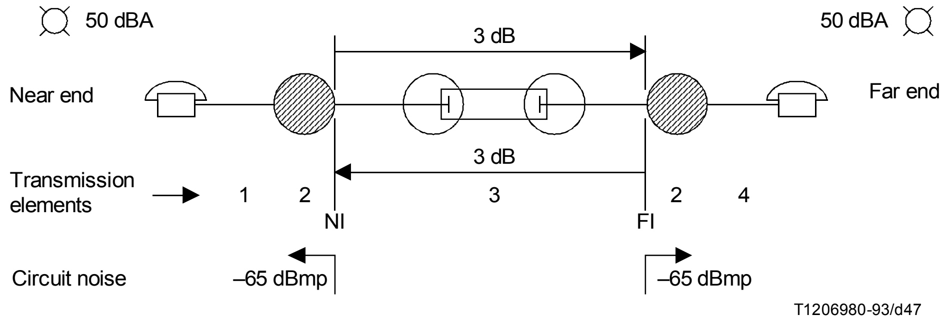


NOTE – These curves show the effect of bandlimiting with ideal lowpass filters.

b) Effect of bandlimiting (lowpass) and loss on the listening effort and conversation opinion scores

FIGURE B.6

HRC 6 – Effect of injected circuit noise level and bandlimiting



Transmission elements

Telephone instruments are BT Type No. 706

1 Line: random selection from a sample of 1800 existing customers lines

2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)

3 PCM system 600 Ω, 3 dB

4 Line: random selection of a line from the same sample of 1800 as in 1 above

			Line 43 (1.3 km)					Line 121 (0.9 km)			
Near end			*IL*	=	64	*IL*	=	64	*Far end*		
STMR	=	10.38	SLR	=	5.45	OLR	=	4.77	RLR	=	–4.44
RLR	=	–4.46	OLR	=	4.77	SLR	=	5.47	STMR	=	8.59
Y_{LE}	=	3.56	V_L	=	–19.07	V_L	=	–19.11	Y_{LE}	=	3.57
Y_C	=	3.06	V_C	=	–22.96	V_C	=	–23.56	Y_C	=	3.07
RN	=	50.00	ICN	=	–65.00	ICN	=	–65.00	RN	=	50.00

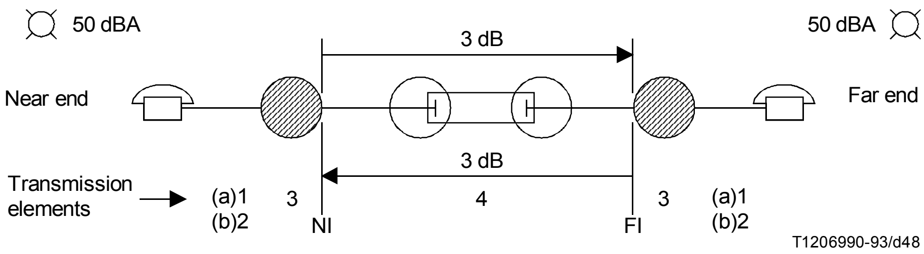
			Line 731 (0.3 km)					Line 87 (0.5 km)			
Near end			*IL*	=	75	*IL*	=	64	*Far end*		
STMR	=	7.19	SLR	=	4.08	OLR	=	2.42	RLR	=	–5.41
RLR	=	–5.74	OLR	=	2.53	SLR	=	4.50	STMR	=	6.77
Y_{LE}	=	3.45	V_L	=	–17.46	V_L	=	–18.16	Y_{LE}	=	3.50
Y_C	=	3.05	V_C	=	–22.29	V_C	=	–23.13	Y_C	=	3.06
RN	=	50.00	ICN	=	–65.00	ICN	=	–65.00	RN	=	50.00

			Line 4 (2.0 km)					Line 776 (0.9 km)			
Near end			*IL*	=	50	*IL*	=	75	*Far end*		
STMR	=	4.33	SLR	=	4.05	OLR	=	2.45	RLR	=	–5.38
RLR	=	–4.65	OLR	=	3.54	SLR	=	4.45	STMR	=	7.28
Y_{LE}	=	3.53	V_L	=	–17.84	V_L	=	–17.83	Y_{LE}	=	3.47
Y_C	=	3.05	V_C	=	–23.53	V_C	=	–22.60	Y_C	=	3.03
RN	=	50.00	ICN	=	–65.00	ICN	=	–65.00	RN	=	50.00

			Line 1018 (2.2 km)					Line 1647 (2.5 km)			
Near end			*IL*	=	50	*IL*	=	40	*Far end*		
STMR	=	8.95	SLR	=	3.41	OLR	=	4.37	RLR	=	–2.72
RLR	=	–5.27	OLR	=	4.59	SLR	=	6.18	STMR	=	8.94
Y_{LE}	=	3.54	V_L	=	–17.17	V_L	=	–19.54	Y_{LE}	=	3.56
Y_C	=	3.03	V_C	=	–21.42	V_C	=	–23.88	Y_C	=	3.07
RN	=	50.00	ICN	=	–65.00	ICN	=	–65.00	RN	=	50.00

FIGURE B.7

HRC 7 – Example with random selection of customers lines



Transmission elements

Telephone instruments are BT Type No. 706, with the values of Z_{SO} modified as required

1 Unloaded cable: as specified below
2 Line: random selection from a sample of 1800 existing customers lines
3 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
4 PCM system 600 Ω, 3 dB

NOTE – See also Tables B.1 and B.2.

FIGURE B.8

HRC 8 – Example of the use of CATNAP in design

TABLE B.1

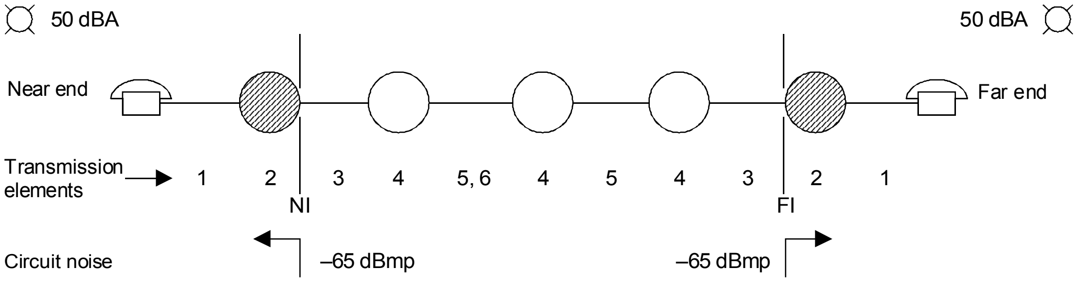
Values of STMR (dB) for specified lines (copper conductors)

Z_{so}	1.6 km 0.5 mm (median)	6 km 0.5 mm	3.7 km 0.4 mm	7.2 km 0.63 mm	10 km 0.9 mm
		<--------------------------(Limiting)-------------------------->			
706	9.9	15.7	7.2	7.5	0.0
Conjugate of input Z	1.8	1.1	0.6	–0.2	–0.6
600 Ω	6.6	–0.8	–1.2	–2.0	–3.0
Suggested values	10.2	13.4	13.8	4.4	–1.3

TABLE B.2

Distribution of STMR for a sample of 40 lines for a Standard 706 and the suggested values of Z_{so}

Z_{so}	Mean	Standard deviation	Maximum value	Minimum value
706	8.3	± 2.5	14.1	3.8
Suggested values	9.4	± 3.1	17.9	4.2



Transmission elements

Telephone instruments are BT Type No. 706

1 Unloaded cable 0-10 km of 0.63 mm (109 ohms/km, 50 nF/km)
2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
3 Loaded junction 19.6 km of 0.9 mm, 88 mH at 1.83 km
4 Transformer feed bridge (50 V)
5 Attenuation 3.5 dB, frequency independent, 600 Ω
6 Channel filtering 300 Hz-3.4 Hz, 600 Ω

NOTE – The results are shown in the curves.

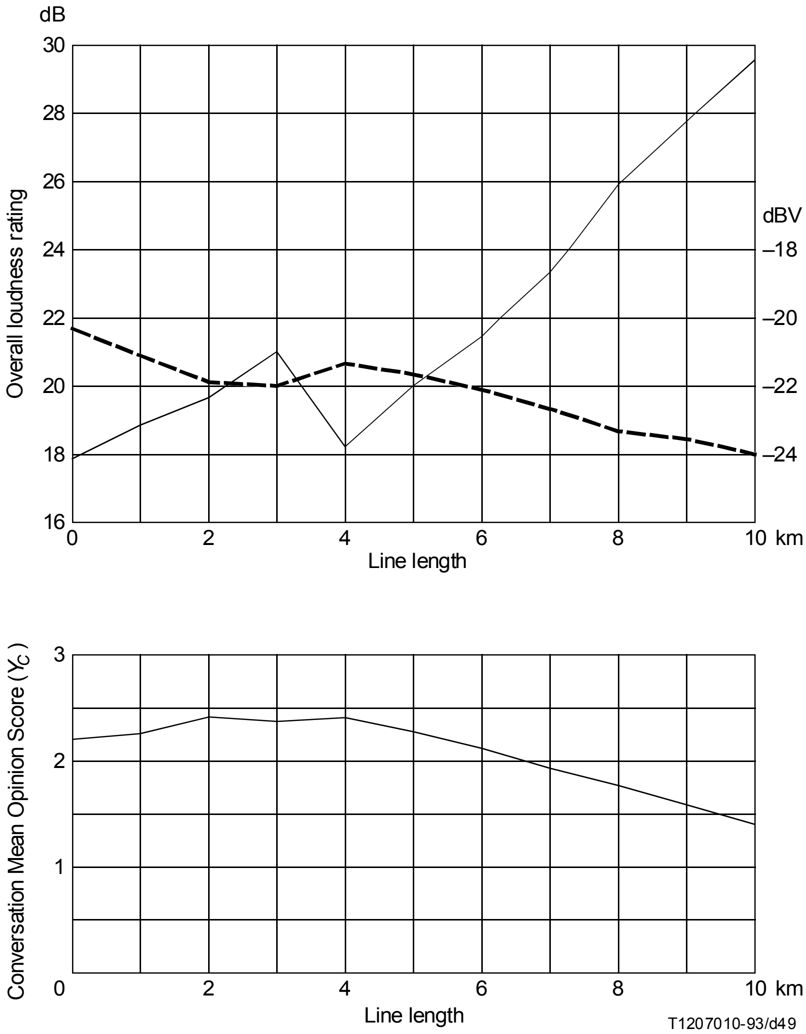


FIGURE B.9

HRC 9 – Effect of varying line length

Annex C

(reference to 4.3.1.2)

Noise spectrum calculation

Expansion from the scalar value of noise to the spectrum values of both room noise and circuit noise is necessary (see Figure 4-4). The spectrum value database of R_N (60 dBA) and V_c (–56.0 dBmp) is shown in Table C.1. The value of room noise is taken from Figure 2/P.45 [40] and Figure A.1/P.80. V_c is a mixture of circuit noise and switching office noise. They are expressed by flat noise and –8 dB/octave noise, respectively. If only a scalar noise level is known as a test condition, and its spectrum value is not known, then a mixed noise spectrum is used in OPINE in which –8 dB octave noise is 10 dB lower than flat noise. Moreover, SRAEN characteristics are added to the flat noise characteristics.

TABLE C.1

Noise spectrum value used in OPINE

R_N = 60 dBA V_C = –56.0 dBmp

No.	Frequency	B_{RNi}	V_{flat} + SRAEN	$V_{-8/oct}$	V_{CQi} = V_{flat} (+) $V_{-8/oct}$
	(Hz)	(dB) 20 µPa/Hz	(dBV/Hz)	(dBV/Hz)	(dBV/Hz)
1	100	42.07	–112.91	–75.25	–75.25
2	125	40.67	–102.61	–77.95	–77.93
3	160	39.07	–98.11	–80.55	–80.47
4	200	37.37	–96.81	–83.25	–83.06
5	250	35.87	–95.21	–85.95	–85.46
6	315	34.37	–93.31	–88.55	–87.29
7	400	32.87	–92.41	–91.25	–88.78
8	500	31.17	–91.91	–93.85	–89.76
9	630	29.57	–91.51	–96.55	–90.32
10	800	27.87	–91.21	–99.25	–90.57
11	1000	26.37	–91.21	–101.95	–90.86
12	1250	24.77	–91.21	–104.55	–91.01
13	1600	23.07	–91.11	–107.25	–91.00
14	2000	21.37	–91.01	–109.95	–90.95
15	2500	19.57	–91.01	–112.55	–90.98
16	3150	17.37	–91.21	–115.25	–91.19
17	4000	14.87	–178.71	–117.95	–117.95
18	5000	12.17	–291.21	–120.55	–120.55
19	6300	9.37	–291.21	–123.25	–123.25
			–56.4 dBmp	–66.4 dBmp	

Annex D

(reference to 4.3.3)

MOS calculation examples

The test condition with an NTT 600 type telephone and a 0.4 mm, 7 dB line as a local telephone circuit (LTC) is considered here. Input data concerning the LTC is shown in Table D.1. In this connection, SLR = 6.6 dB, and RLR = –3.8 dB. The test conditions and calculated results for fundamental factors are shown in Table D.2.

The output of the overall electro-acoustic calculation (see 4.3.1) for test condition No. 11 in Table D.2 is shown in Figure D.1, where OLR is 6.4 dB.

TABLE D.1

Local telephone circuit sensitivity
(NTT 600-type telephone set with a 0.4 mm, 7 dB line)

No.	Frequency	S_{MJi}	S_{JEi}	L_{MESTi}	L_{RNSTi}
	(Hz)	(dB) rel V/Pa	(dB) rel Pa/V	(dB)	(dB)
1	100	–22.3	–40.0	5.3	28.6
2	125	–25.1	–2.7	6.7	26.3
3	160	–23.8	2.5	5.0	20.8
4	200	–18.8	7.3	2.3	14.1
5	250	–14.4	11.3	–3.0	5.6
6	315	–12.3	14.6	–6.4	–1.3
7	400	–12.5	15.9	–5.6	–1.8
8	500	–12.6	15.7	–3.6	–0.3
9	630	–12.3	14.9	–2.1	2.8
10	800	–11.9	14.4	–0.4	3.9
11	1000	–11.6	14.5	0.1	3.4
12	1250	–12.0	14.8	0.0	3.1
13	1600	–12.0	14.1	0.1	0.1
14	2000	–9.8	14.4	–3.3	–2.1
15	2500	–10.0	16.2	–5.0	3.4
16	3150	–11.0	11.5	2.7	15.0
17	4000	–16.8	8.9	11.1	22.3
18	5000	–27.9	–30.0	28.1	35.1
19	6300	–32.0	–30.0	32.7	35.3

TABLE D.2

Example of estimated results for fundamental factors by OPINE

Test conditions (STMR = 7.1 dB)						Conversion to OPINE input			Output					
No.	Noise OLR (dB)	R_N (dBA)	Circuit noise (dBmp)	Switching noise (dBmp)	Frequency charac-teristic (Table D.3)	OLR (dB)	L (dB)	V_C (dBmp)	PI_{EL}	PI_N	PI_{AD}	PI_{ST}	OPI	MOS
1	–3.8	0			1	–3.6	–7.3	–95.1	0.63	0.00	0.19	0.15	0.97	2.58
2	1.2	0			1	1.4	–2.3	–95.1	0.23	0.00	0.10	0.15	0.49	3.04
3	6.2	0			1	6.4	2.7	–95.1	0.03	0.00	0.09	0.15	0.27	3.23
4	11.2	0			1	11.4	7.7	–95.1	0.40	0.00	0.12	0.15	0.67	2.88
5	16.2	0			1	16.4	12.7	–95.1	0.80	0.00	0.08	0.15	1.03	2.52
6	21.2	0			1	21.4	17.7	–95.1	1.20	0.00	0.04	0.15	1.40	2.16
7	26.2	0			1	26.4	22.7	–95.1	1.61	0.00	0.04	0.15	1.81	1.75
8	31.2	0			1	31.4	27.7	–95.1	2.02	0.00	0.02	0.15	2.20	1.37
9	–3.8	60	–56.9	–62.2	1	–3.6	–7.3	–55.8	0.56	0.21	0.19	0.15	1.12	2.44
10	1.2	60	–56.9	–62.2	1	1.4	–2.3	–55.8	0.14	0.21	0.10	0.15	0.61	2.93
11	6.2	60	–56.9	–62.2	1	6.4	2.7	–55.8	0.15	0.21	0.09	0.15	0.60	2.94
12	11.2	60	–56.9	–62.2	1	11.4	7.7	–55.8	0.60	0.21	0.12	0.15	1.08	2.48
13	16.2	60	–56.9	–62.2	1	16.4	12.7	–55.8	1.09	0.21	0.08	0.15	1.54	2.02
14	21.2	60	–56.9	–62.2	1	21.4	17.7	–55.8	1.62	0.21	0.04	0.15	2.03	1.53
15	26.2	60	–56.9	–62.2	1	26.4	22.7	–55.8	2.21	0.23	0.04	0.15	2.64	0.95
16	31.2	60	–56.9	–62.2	1	31.4	27.7	–55.8	2.87	0.26	0.02	0.15	3.30	0.41
17	1.2	60	–56.9		1	1.4	–2.3	–57.0	0.15	0.16	0.10	0.15	0.57	2.97
18	11.2	60	–56.9		1	11.4	7.7	–57.0	0.59	0.16	0.12	0.15	1.02	2.53
19	21.2	60	–56.9		1	21.4	17.7	–57.0	1.61	0.16	0.04	0.15	1.96	1.60
20	31.2	60	–56.9		1	31.4	27.7	–57.0	2.84	0.21	0.02	0.15	3.23	0.47
21	1.2	50	–56.9	–62.2	1	1.4	–2.3	–55.8	0.17	0.21	0.10	0.15	0.64	2.90
22	11.2	50	–56.9	–62.2	1	11.4	7.7	–55.8	0.53	0.21	0.12	0.15	1.01	2.54
23	21.2	50	–56.9	–62.2	1	21.4	17.7	–55.8	1.48	0.21	0.04	0.15	1.89	1.67
24	31.2	50	–56.9	–62.2	1	31.4	27.7	–55.8	2.59	0.22	0.02	0.15	2.99	0.65
25	1.2	45	–68.2	–68.2	1	1.4	–2.3	–65.2	0.20	0.02	0.10	0.15	0.48	3.05
26	13.2	45	–68.2	–68.2	1	13.4	9.7	–65.2	0.63	0.02	0.12	0.15	0.92	2.63
27	26.2	45	–68.2	–68.2	1	26.4	22.7	–65.2	1.80	0.02	0.04	0.15	2.02	1.55
28	1.2	45	–63.8	–68.2	1	1.4	–2.3	–62.5	0.20	0.04	0.10	0.15	0.50	3.03
29	13.2	45	–63.8	–68.2	1	13.4	9.7	–62.5	0.65	0.04	0.12	0.15	0.96	2.60
30	26.2	45	–63.8	–68.2	1	26.4	22.7	–62.5	1.84	0.04	0.04	0.15	2.07	1.49
31	2.2	60	–56.9	–62.2	3	2.5	–2.4	–55.8	0.07	0.21	0.28	0.15	0.72	2.83
32	12.2	60	–56.9	–62.2	3	12.5	7.6	–55.8	0.69	0.21	0.20	0.15	1.25	2.30
33	22.2	60	–56.9	–62.2	3	22.5	17.6	–55.8	1.71	0.21	0.12	0.15	2.19	1.37
34	32.2	60	–56.9	–62.2	3	32.5	27.6	–55.8	2.95	0.26	0.04	0.15	3.41	0.35
35	4.1	60	–56.9	–62.2	7	5.1	–2.3	–55.8	0.02	0.21	0.45	0.15	0.84	2.71
36	14.1	60	–56.9	–62.2	7	15.1	7.7	–55.8	0.89	0.21	0.31	0.15	1.57	1.99
37	24.1	60	–56.9	–62.2	7	25.1	17.7	–55.8	1.92	0.22	0.18	0.15	2.47	1.10
38	34.1	60	–56.9	–62.2	7	35.1	27.7	–55.8	3.16	0.27	0.06	0.15	2.64	0.23

TABLE D.3

Attenuation/frequency characteristics used in Table D.2

Frequency	1	2	3
(Hz)	SRAEN (dB)	(Note 1) (dB)	(Note 2) (dB)
100	21.7	40.0	76.0
125	11.4	32.0	60.0
160	6.9	23.0	47.0
200	5.6	17.2	36.0
250	4.0	12.0	24.5
315	2.1	6.5	15.0
400	1.2	2.5	7.0
500	0.7	1.0	2.5
630	0.3	0.5	0.5
800	0.0	0.0	0.0
1000	0.0	–0.1	0.0
1250	0.0	–0.1	0.0
1600	–0.1	–0.3	0.2
2000	–0.2	–0.1	0.9
2500	–0.2	0.5	2.5
3150	0.0	4.0	9.0
4000	87.5	12.5	19.5
5000	200.0	22.0	30.0
6300	200.0	32.0	41.0

NOTES

1 Three 4-wire circuit chains. 50% limit characteristics.

2 Seven 4-wire circuit chains. 95.5% limit characteristics.

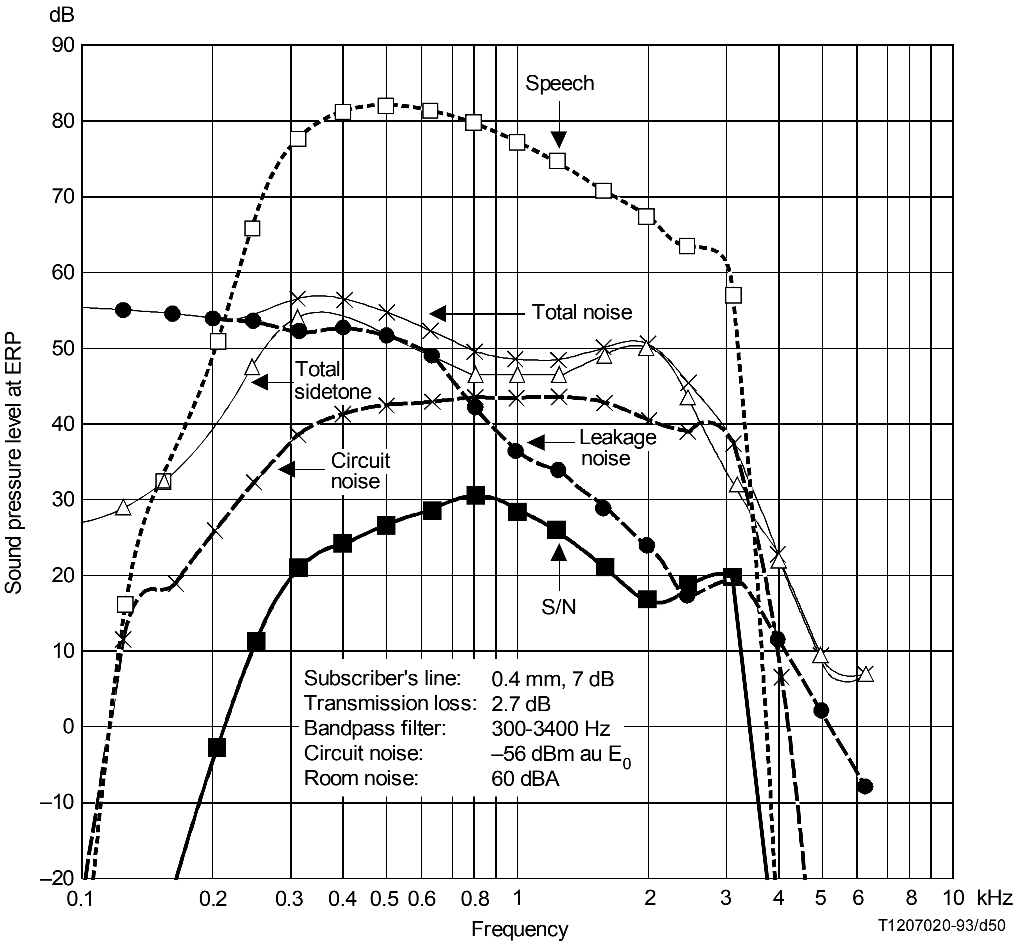


FIGURE D.1

Speech and noise level at ERP

Annex E

(reference to 4.3.2.2)

Derivation of equation (4-16)

From equations (4-9) and (4-10) of Recommendation P.79,

$$OLR = \overline{L_{ME}} - \overline{L_{RNE}} = -\frac{1}{m} 10 \log_{10} \sum_{i=1}^{M} 10^{\frac{-mL_{MEi}}{10}} G_i \Delta f_i - \left(-\frac{1}{m}\right) 10 \log_{10} \sum_{i=1}^{M} 10^{\frac{-mL_{RMEi}}{10}} G_i \Delta f_i \qquad \text{(E-1)}$$

Taking the logarithm of equation (4-12),

$$10 \log_{10} \lambda_E = 10 \log_{10} C + 10 \log_{10}\left[\sum_{i=1}^{M} 10^{\frac{-mb_n}{10}} \cdot 10^{\frac{-mL_{MEi}}{10}} G_i \Delta f_i\right] = K - mb_n - m\overline{L_{ME}} \qquad \text{(E-2)}$$

Similarly,

$$10 \log_{10} \lambda_0 = 10 \log_{10} C + 10 \log_{10}\left[\sum_{i=1}^{M} 10^{\frac{-mL_{\Phi MEi}}{10}} G_i \Delta f_i\right] = K - m\overline{L_{\Phi ME}} \qquad \text{(E-3)}$$

where $L_{\Phi MEi}$ is the loss in dB that gives the optimum loundness when noise is not present.

Substitution of these into equation (4-15), we get

$$PI_{EL} = \sqrt{\frac{C_2}{C_1}\left(10 \log_{10}\frac{\lambda_E}{\lambda_0}\right)^2 + C_2} - \sqrt{C_2} = \sqrt{\frac{C_2}{C_1}\left\{- mb_n - m\overline{L_{ME}} - (-m\overline{L_{\Phi ME}})\right\}^2 + C_2} - \sqrt{C_2}$$

Since $OLR_0 = L_{\Phi ME} - L_{RNE}$, then

$$PI_{EL} = \sqrt{\frac{C_2}{C_1} m^2 (b_n + OLR - OLR_0)^2 + C_2} - \sqrt{C_2} \qquad \text{(E-4)}$$

which is the same as equation (4-16).

In employing equations (4-15) and (4-16), a constant is necessary for each, that is λ_0/C for (4-15) and OLR_0 for (4-16). Adaptation of the values in Table 4-5 allows a 0.004 error for two different PI_{EL} calculations. This error, however, does not cause further errors in subsequent calculations.

Annex F

(reference to 4.3.3)

Psychological evaluation model

This annex gives a detailed derivation of equations (4-34) and (4-35). The model is a complete adaptation of [39].

F.1 Psychological model for evaluation

According to the model in reference [39], an evaluation value for a test condition on a psychological continuum is shown in Figure F.1. p_K is defined on page 10 of the reference, and is the probability of voting K as an opinion score for a test condition. The correspondences of opinion scores to ranges in the psychological continuum are:

Continuum range		*Opinion score*
$-\infty$	0.5	0
0.5	1.5	1
1.5	2.5	2
2.5	3.5	3
3.5	∞	4

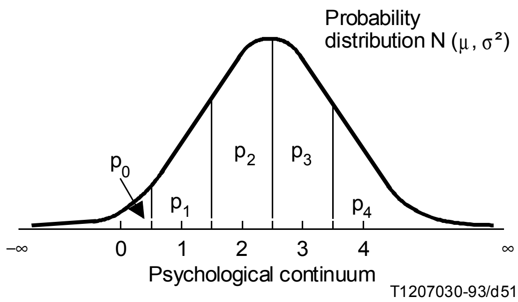


FIGURE F.1

Evaluation distribution for a test condition on a psychological continuum

These assumptions satisfy the following equation:

$$MOS = \sum_{k=0}^{4} k\, p_k \qquad \text{(F-1)}$$

which is the same as equation (4-34).

F.2 Derivation of equation (4-35) from equation (4-34)

The cumulative probability of $N(\mu, \sigma^2)$ is expressed using a standard normal distribution function as follows:

$$p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t \tag{F-2}$$

Using equation (F-2), equation (4-34) is expressed as:

$$MOS = \frac{1}{\sqrt{2\pi}} \left\{ 0 \times \int_{-\infty}^{(0.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t + 1 \times \int_{(0.5-\mu)/\sigma}^{(1.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t + 2 \times \int_{(1.5-\mu)/\sigma}^{(2.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t \right.$$

$$\left. + 3 \times \int_{(2.5-\mu)/\sigma}^{(3.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t + 4 \times \int_{(3.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t \right\} \tag{F-3}$$

By changing the multiplication into a repetition of additions, and by changing the association (combination) of addition, equation (F-3) becomes:

$$MOS = \frac{1}{\sqrt{2\pi}} \left\{ \int_{(0.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t + \int_{(1.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t + \int_{(2.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t \right.$$

$$\left. + \int_{(3.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t \right\} \tag{F-4}$$

Since

$$\frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} \exp(-t^2/2)\mathrm{d}t = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-t^2/2)\mathrm{d}t \tag{F-5}$$

then

$$MOS = 4 - \sum_{k=0}^{3} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(k+0.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t \tag{F-6}$$

Replacement of μ by P results in equation (4-35), which then enables the use of a standard normal distribution table.

Annex G

Objective method of estimating the quality of speech degraded by non-linear distortion

Methods for predicting the subjective quality of transmission systems by objective measurements, which are currently under study are described in this annex. It has not been possible up to now to recommend a single method which is applicable over a wide range of non-linear distortion. An additional method based on a psycho-acoustic approach has been introduced [43].

G.1 Introduction

One of the most popular scales for evaluating the quality of speech transmitted through non-linear devices is the opinion score in which a subjective judgement is made on the actual speech. The mean opinion score (MOS), which is the average of scores given under pertinent conditions, has been used as an index for network planning and for evaluating codec performance. The MOS has higher reliability when the number of talkers and listeners is larger. For practical reasons, however, the number of subjects is restricted.

The previous approach, called subjective measurement, takes a lot of time and effort. As a result, the MOS has been objectively estimated from physical measurements of non-linear distortion by modelling and evaluating the behaviour of listeners, and using an artificial voice which is recommended by Recommendation P.50.

For a PCM codec, one example of the non-linear devices used in modern networks, objective measurement has been applied using either a sinusoidal signal or frequency band limited noise (see Recommendations G.712-G.715). This method is only one application of the conventional techniques, which measures harmonic distortions in analogue circuits, to a PCM codec. However, the method is valid only when the distortion spectrum is flat, as in PCM coding.

The signal-to-noise ratio, SNR, derived from the difference between input and output waveforms of a signal, has also been used in the evaluation as a performance index of the low-bit-rate coding algorithm. However, the contemporary sophisticated coding algorithm statistically exploits redundancies of the human voice. Therefore, a flat distortion spectrum is not expected. The objective measurement methods recommended here are mainly developed for application to such non-linear distortion.

G.2 Scope

The application of the objective method will be restricted to the digital speech codec in the first stage. This implies that input and output signals can be picked up at the input and output ports of the codec. Additional analogue noise or frequency/attenuation distortion generated from the codec is assumed to be negligible. The input signal spectrum can be changed according to the circuit connected to the input side of the codec. System delay is the only problem to be compensated. Application of specific methods should be restricted to this area.

G.3 Measurement signal

Artificial male and female voices, which conform to Recommendation P.50 should be used. If the signal input to the codec is fed from a telephone set, the voices should be shaped according to the sensitivity/frequency characteristics of the set.

G.4 Preprocessing

G.4.1 Time alignment between input and output signal

The time delay between input and output signals is produced by the encoding and decoding processes. This delay causes serious measurement error if time alignment is not applied. One of the most effective methods for measuring the delay is to detect the peak of cross correlation between the input and output signals.

G.4.2 Windowing

A Hamming window is applied. The window length (one frame) is fixed at 256 samples at a 8 kHz sampling rate (32 ms). In most cases, windowing is done without overlapping, the 50% overlapping is effective when coherence function is used.

G.4.3 Segment clustering

G.4.3.1 Method 1 – Segmental power level

The quartiles are produced by developing a cumulative distribution based on segmental power in the source speech and then creating four groups. The four groups are determined by rank ordering the segments according to power level and assigning those segments to clusters based on a percentile ranking. (i.e. 0-25 percentile is assigned to cluster 1, 25-50 percentile is assigned to cluster 2, and so on.)

G.4.3.2 Method 2 – Parameters derived from clusters

Short-term segments in speech signals are classified into several clusters which share properties (parameter values) that makes them sound much the same to humans. Parameter values for a segment are used to assign that segment to a cluster. The parameters used in segmenting clusters are those which are to be for pattern matching. The number of clusters and cluster boundaries are determined by k-Nearest Neighbour (kNN) analysis. This analysis applied to the input speech signal and assignments of segments to a given cluster are made. When segments are assigned to clusters, a record is kept of which segment is assigned to which cluster. This record is used to assign segments to cluster in the output speech signal.

G.4.4 Pre-emphasis

In the method using the LPC technique, signals are emphasized using a first-order differentiator to extract the LPC parameter to improve the accuracy of frequency analysis.

G.4.5 Elimination of pauses

Short silences between passages are eliminated from the measurement. Pauses are detected according to the following rules:

i) If the recorded background noise level is known (and given to the host laboratory), segments having a power level lower than the threshold for more than six successive 32-ms frames are judged to be pauses. The threshold is set above the noise level.

ii) If the recorded background noise level is not known (or not provided), it is estimated from a period of silence on the recorded tape or from idle noise generated by the codec.

G.4.6 Amplitude normalization

Some types of non-linear distortion produce amplitude differences between I/P and O/P signals. If some measures apply to such non-linear distortion, their individual values may be mainly determined by the corresponding amplitude difference rather than the non-linear distortion. In this case, the O/P amplitude signal should be adjusted so as to give the same level as the I/P signal.

G.5 Objective parameters

The objective speech quality is estimated on the basis of a single or multiple parameters specifying the speech signal in the time or frequency domain. The parameters are also closely related to subjective scores. The parameters can be classified into "base parameters" and "distortion parameters". Base parameters specify inherent physical characteristics only for input or output speech. The distortion parameters, on the other hand, relate the differences between input and output speech signals. Most objective measurement methods have been suggested on the basis of this distortion measure. Although there are a number of distortion parameters, those which are conventionally and widely used for the distortion index and those which are necessary in the estimation processes are recommended.

G.5.1 Method 1 – Signal-to-noise ratio (SNR)

Equation (G-1) gives the conventional signal-to-quantizing-noise ratio, Qt, which has been used as a common objective quality measure.

$$Qt = 10 \log_{10} \frac{\sum_{j=1}^{M} x(j)^2}{\sum_{j=1}^{M} \{x(j) - y(j)\}^2} \qquad \text{(G-1)}$$

Here, $x(j)$ and $y(j)$ are the input and output signals for the codec, and M is the number of samples in the input part. The alternative measure derived from Qt is segmental SNR, Qt_{seg}, which is defined by equation (G-2).

$$Qt_{seg} = \frac{1}{N} \sum_{i=1}^{N} Qt_i \qquad \text{(G-2)}$$

Here, N is the number of signal frames and Qt_i is the signal-to-noise ratio in the signal frame i.

Modified definitions for SNR and SNR_{seg} are derived from the power spectrum. $Sx_i(j)$, and $Sy_i(j)$ are the power spectra of the input and output signals of the j-th frequency at the i-th segment. The defined signal-to-noise ratios in the frequency domain, Qf and Qf_{seg} are:

$$Qf = 10 \log_{10} \frac{\sum_{j=1}^{M} Sx(j)}{\sum_{j=1}^{M} Sx(j) - \sum_{j=1}^{M} Sy(j)} \qquad \text{(G-3)}$$

$$Qf_{seg} = \frac{1}{N} \sum_{i=1}^{N} Qf_i \qquad \text{(G-4)}$$

G.5.2 Method 2 – Cepstrum distance

The LPC Cepstrum Distance is approximated using cepstral coefficients as follows:

$$CD = \left[2\Sigma \{C_x(i) - C_y(i)\}^2 \right]^{1/2} \quad \text{(dB)} \qquad \text{(G-5)}$$

where $C_x(i)$ and $C_y(i)$ are cepstral coefficients of the input and output signals. This measure relates to the spectrum envelope difference between the source and are processed speech.

G.5.3 Method 3 – Coherence function

Coherence function γ is defined by a cross correlation function and an auto correlation function for input and output speech samples, x_i and y_i.

$$\gamma^2 = \frac{|G_{ab}|^2}{G_{aa} \times G_{bb}} \tag{G-6}$$

where G_{aa}, G_{bb}, and G_{ab} are input auto spectrum, output spectrum, and cross spectrum, respectively.

In equation form,

$$G_{ab} = \sum_{n=1}^{N} X_n (f)\, Y^*{}_n \tag{G-7}$$

$$G_{aa} = \sum_{n=1}^{N} X(f)^2 \tag{G-8}$$

$$G_{bb} = \sum_{n=1}^{N} Y(f) \tag{G-9}$$

$$\delta = \frac{1}{NPf_s} \tag{G-10}$$

where N is the number of segments in the source signal, P is the number of samples in each segment, and f_s is the sampling frequency.

Using γ and $Y(f)$, coherent power spectrum, $CP(f)$, and residual non-coherent power, $NCP(f)$, are calculated as follows:

$$CP(f) = \gamma^2 (f) \cdot \delta \cdot G_{bb} \tag{G-11}$$

$$NCP(f) = \left[1 - \gamma^2 (f)\right] \cdot \delta \cdot G_{bb} \tag{G-12}$$

G.5.4 Method 4 – Information index

The information index is based on the concept of mutual information defined by Shannon. The principle of this method is described in [24] and its implementation in 3.2. First, a segmental SNR is calculated for frequency bands having the same articulation importance. The SNRs are then transformed to Q-values. The information index, II, is calculated by the following equations:

$$\mathrm{II} = \sum_{j=1}^{16} B_j \cdot V_j \tag{G-13}$$

with

$$V_j = 3 \Big/ \left(0{,}1 + 10^{-(Qf_j + C_j)/10}\right) \tag{G-14}$$

where B_j is the frequency weight allocated to the j-th frequency band, and Qf_j is the average of q_j over all frames. The Qf_j is the equivalent SNR expressed in MNRU, and C_j is a correction factor for the critical bandwidth in the hearing mechanism. C_j and B_j are provided in Table G.1. For a codec the sum $Qf_j + C_j$ is approximately given, using Qf, by the following formulae: where $d = Qt_{seg} - Qt$ for each particular type of codec. Examples of d for typical codecs are shown in Table G.2.

$$Qf_j + C_j = Qf + C_j + d \cdot \tanh[0.0798\,(Qf + C_j) - 0.356] \text{ for } Qf + C_j < -3.57 \tag{G-15}$$

$$\begin{aligned} Qf_j + C_j = {} & 4.34 \ln[\exp\{0.23026\,(Qf + C_j + 5.15)\} - 1] \\ & + d\,[0.276\,(Qf + C_j) + 0.3859] \qquad \text{for } -3.57 < Qf + C_j < 0 \end{aligned} \tag{G-16}$$

$$\begin{aligned} Qf_j + C_j = {} & 4.3429 \cdot \ln[\exp\{0.23026\,(Qf + C_j + 5.15)\} - 1] \\ & + d \cdot \tanh[0.062715\,(Qf + C_j) + 0.310925] \qquad \text{for } Qf + C_j > 0 \end{aligned} \tag{G-17}$$

TABLE G.1

j	Frequency (Hz)	B_j	C_j
1	100	0.00804	0
2	125	0.01042	1.25
3	160	0.0138	2.0
4	200	0.01788	2.6
5	250	0.02392	3.5
6	315	0.03246	4.9
7	400	0.04471	5.8
8	500	0.05981	6.35
9	630	0.07789	7.25
10	800	0.0839	7.35
11	1000	0.0899	7.8
12	1250	0.09627	8.05
13	1600	0.10376	8.25
14	2000	0.11097	8.3
15	2500	0.11859	8.18
16	3150	0.12694	7.95
17	4000	0.13607	7.57
18	5000	0.14506	7.25
19	6300	0.15487	7.2
20	8000	0.16554	6.8

TABLE G.2

Type of codec	Bit rate (kbit/s)	d
PCM	64	0
ADPCM with fixed predicter	16	–2.8
	24	–2.3
	32	–2.4
ADPCM with variable predicter	16	1.7
	24	2.1
	32	2.6
Sub-band coding	24	–2.1
	32	–1.0

G.6 Estimation process

G.6.1 Method 1 – Using regression equation

For *CD*,

$$MOS = a \cdot CD^2 - b \cdot CD + c \qquad \text{(G-18)}$$

The parameters a, b and c are determined from the opinion score. The values derived for a, b and c from a Japanese experiment were 0.0415, 0.8010 and 3.5620 respectively.

For II,

$$x = \ln\left[\frac{\mathrm{II} - \mathrm{I}_{min}}{\mathrm{I}_{max} - \mathrm{II}}\right] \qquad y = \ln\left[\frac{Y - Y_{min}}{Y_{max} - Y}\right] \qquad \text{(G-19)}$$

$$y = A \cdot x - B \qquad \text{(G-20)}$$

where Y_{max} (Y_{min}) is the maximum (minimum) score in an extended series of opinion tests and I_{max} (I_{min}) is the corresponding value of II; these can be directly obtained in the tests or extrapolated. The values derived from French experiment, I_{max}, I_{min}, Y_{max}, and Y_{min} are 27.6, 0, 3.4 and 0 (on a 0-4 scale), respectively for various codecs and MNRUs. The values derived from BNR (Canada) tests for input-output non-linearities are 25, 9.89, 4.2 and 1.057 (on a 1-5 scale) respectively.

G.6.2 Using opinion model

G.6.2.1 Extended Richards model

The first step for normalizing the coherent power level ("signal level") to the "preferred level" (i.e. 82 dBspl) is explained. Using power addition the non-coherent power spectrum ("Noise spectrum") is combined with the hearing threshold for continuous spectrum sound to form a new masking noise spectrum (MNS). The hearing threshold is given in Table G.3. From that, the sensation level *Z* is found by subtracting non-coherent power from coherent power, making an additive index *P*(*Z*) using the modified growth functions.

$$Z < A \; P(z) = 10^{(Z + B)/10} \qquad \text{(G-21)}$$

$$Z \geq A \; P(Z) = 1 - \left[10^{(Z + C)/10}\right]^D \qquad \text{(G-22)}$$

where

$$A = 2.792, \quad B = -6.646, \quad C = 0.5 \text{ et } D = -0.7$$

The product (or sum if expressed in decibels) of *P*(*Z*) and the frequency weighting factor B' (10 $\log_{10} B'$ is given in Table G.3) is then integrated over the relevant frequency range to obtain the listening opinion index, LOI. (The values of $\beta_0 - K$ and B_1 are shown in Table G.3.)

TABLE G.3

Frequency f(Hz)	Hearing threshold $\beta_0 - K$	Frequency weighting $10 \log B'$
100	+17.5	–35.8
200	5.0	–34.2
300	0.0	–33.3
400	–3.0	–32.9
500	–5.0	–32.9
600	–6.0	–33.0
800	–8.0	–33.5
1000	–9.0	–34.0
1250	–8.5	–34.7
1600	–8.0	–35.7
2000	–9.0	–37.3
2500	–11.5	–39.4
3000	–14.0	–41.3
3500	–13.5	–42.9
4000	–13.0	–44.0
5000	–12.5	–45.5
6000	–11.5	–46.7
8000	–9.0	–48.2

In practice, assuming that the sensation level Z is approximately constant within suitable chosen narrow frequency bands Δf (frequency resolution determined by the length of the segments used in FFT calculations), the integration is replaced by a summation of products:

$$B' P(z) \Delta f \qquad \text{(G-23)}$$

Up to this point all calculations were done separately for each individual quartile, thus yielding four values of listening opinion indices, LOI. Since these indices are assumed to be additive they can be averaged using the following weighting factors:

Lowest level	Q1	0.19
	Q2	0.21
	Q3	0.53
Highest level	Q4	0.07

The weighting factors are applied to compensate for the fact that the simulated speech signal as described in [41] has a segmented power level distributed over a much narrower range than those of the real speech. Further modification of the simulated speech signal may eventually eliminate the need for applying a different weighting to each quartile.

The final LOI may be transformed into the mean opinion score, MOS, using the following modified relationship.

$$MOS = \frac{1 + 5e^x}{1 + e^x} \qquad \text{(G-24)}$$

where

$$x = E \cdot \ln \frac{LOI}{LOI_{lim} - LOI} + F \qquad \text{(G-25)}$$

$$E = 1.145, \quad F = -1.195 \quad \text{et} \quad LOI_{lim} = 0.885$$

G.6.2.2 OPINE model using equivalent *Q*

An OPINE description related to quantizing distortion is extracted from a relevant section in the Handbook on Telephonometery.

At first, CD is converted to an equivalent Q using the following regression equation:

$$Q = 0.49\,CD^2 - 8.425\,CD + 42.856 \quad \text{(dB)} \tag{G-26}$$

The opinion equivalent white noise level of quantizing distortion is as follows:

$$V_{wop} = -2.022\,PI_q^{\,3} - 7.51\,PI_q^{\,2} + 21.9\,PI_q - 76.9 - (OLR - 7.5) \tag{G-27}$$

$$PI_q = -0.0000218\,Q_{op}^{\,3} + 0.00489\,Q_{op}^{\,2} - 0.283\,Q_{op} + 4.915 \tag{G-28}$$

or in a simpler form with SNR_W:

$$V_{wop} = -26.9 - SNR_w - (OLR - 7.5) \tag{G-29}$$

$$SNR_w = -0.0467\,Q_{op}^{\,2} + 3.632\,Q_{op} - 21.51 \tag{G-30}$$

$$V_{cq} = V_c\,(+)\,V_{wop} \tag{G-31}$$

where

V_{wop}	is the opinion (PI) equivalent white noise level at the input to the receiving local telephone circuit. (RLR = –4.0 dB),
PI_q	is the Performance Index for quantizing distortion,
Q_{op}	is the opinion equivalent speech to the speech correlated noise ratio (dB),
SNR_W	is the opinion equivalent speech to the white noise ratio (dB),
V_c	is the circuit noise level at the input to the receiving end (dBmp), and
V_{cq}	is the equivalent circuit noise level when both circuit noise and quantizing distortion are present.

In the next step, V_{cq} is converted to the performance index of noisiness, PI_N. The noise has two different effects on subjects according to the speech environment. One is noise in speech intervals and the other is noise in non-speech intervals.

In speech intervals, loudness with a lower signal-to-noise ratio, S/N, is evaluated as worse than an equal loudness with a higher S/N. Using the total noise, N_i [Sound pressure level (dB)] for band i at ERP, the performance index for noisiness, PI_N is expressed as the sum of PI_{IN} (PI for idle circuit noise) and PI_{SN} (PI for speech interval noise).

PI_{IN} can be expressed as a power function weighted by an A-curve with respect to the threshold of N_i.

Here

$$N_i = \begin{cases} N_i - N_{\text{th}}, & N_i \geq N_{\text{th}} \\ 0\quad, & N_i < N_{\text{th}} \end{cases} \tag{G-32}$$

$$PI_{IN} = C_3 \sum_{i=1}^{M} \left\{ 10^{A_i/10} \left(10^{nN'_i/10} - 1\right) \right\} \tag{G-33}$$

where

N'_i	level above the noise threshold,
N_{th}	noise threshold [sound pressure level (dB)],
n	exponent, and
A_i	A-weight at frequency band i (dB).

It is necessary to estimate n and N_{th}.

PI_{SN} is evaluated by the following linear formula, where SNR is S/N and SNR_{th} is the threshold below which S/N has no effect on the evaluation.

$$SNR = 10 \log_{10} \left(\sum_{i=1}^{M} 10^{S_i/10} \Big/ \sum_{i=1}^{M} 10^{N_i/10} \right) \quad \text{(G-34)}$$

$$PI_{SN} = \begin{cases} C_4 \ (SNR - SNR_{th}), & SNR \leq SNR_{th} \\ 0 \qquad\qquad\qquad\quad , & SNR > SNR_{th} \end{cases} \quad \text{(G-35)}$$

Here, S_i is the sound spectrum of band i at ERP [sound pressure level (dB)].

$$PI_N = PI_{IN} + PI_{SN} \quad \text{(G-36)}$$

Overall performance index, OPI, can be derived as the sum of the performance indices for other factors.

$$OPI = \sum_{I=1}^{L} PI_I \quad \text{(G-37)}$$

The evaluation results under the same test conditions for a large number of subjects are assumed to obey a normal distribution on a psychological scale. Let P be the mean overall evaluation on this psychological scale, P can then be given as:

$$P = P_0 - OPI \quad \text{(G-38)}$$

where P_0 represents P with no degradation.

Both overall evaluation and overall degradation obey normal distributions with variance σ^2 and means P and OPI, respectively, for given test conditions. In the opinion test, five categories are used. Assuming that the intervals between categories are equal, the MOS can be evaluated as:

$$MOS = \sum_{k=0}^{4} k \, p_k \quad \text{(G-39)}$$

where p_k is the ratio of evaluation category k to all the categories. In the actual situation, the following equation is used

$$MOS = 4 - \sum_{k=0}^{3} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(k+0.5-P)/\sigma} \exp(-t^2/2) \mathrm{d}t \quad \text{(G-40)}$$

G.6.3 Method 3 – Using pattern matching

The method is applied in two stages, in the training stage, given by (G-41) and (G-42) below, speech records that have been subjectively tested are analysed to extract "training statistics". These speech records make up a training data base that includes speech from a variety of relevant codecs. In the testing stage, given by equations (G-43) through (G-45) below, a speech record of unknown quality is analysed and assigned a quality score based on previously measured training statistics.

The technique is based on simple Bayes' probability rules and requires the estimation of basic probability density function (pdf). This step is referred to as "training" because it adopts the system to a particular application.

The conditional probability density function of x_i for the m-th distortion can be estimated using the k-nearest neighbour method:

$$p(x_i \mid d_m) = (k - 1) / (N \cdot \upsilon(x_i)) \tag{G-41}$$

where i is the frame number, x_i is a vector of parameter measurements, N is the total number of frames per distortion, and $\upsilon(x_i)$ is the volume of a hypersphere with a radius equal to the distance from x_i to the k-th nearest vector belonging to distortion d_m. Estimating the conditional probability density function (cpdf) for each distortion consists of storing the training parameter values in memory and computing them.

An alternative estimate of $p(x_i|d_m)$ can be made by modelling the density as a Gaussian mixture. K-means cluster analysis is used to identify clustering in the parameter data for each distortion. The output of the cluster analysis consists of mean vector xm_{mc} and covariance matrix C_{mc} for the m-th distortion and c-th cluster. The Gaussian mixture cpdf estimate is formed by fitting a Gaussian function to each cluster and forming a weighted sum of these functions:

$$p(x_i \mid d_m) = \sum_{c=i}^{N_{mc}} \frac{NC_m}{N_m} \frac{1}{(2\pi)^{p/2} \mid C_{mc} \mid^{1/2}} \exp\left\{-\frac{1}{2}(x_i - xm_{mc})^T C_{mc}^{-1} (x_i - xm_{mc})\right\} \tag{G-42}$$

where p is the number of parameters (dimensions) in vector x_i, N_{mc} is the number of vectors assigned to the c-th cluster of the m-th distortion, NC_m is the number of clusters in the m-th distortion and N_m is the total number of training vectors in the m-th distortion. Typically, (G-41) is used during feature evaluation and selection, while the Gaussian mixture (G-42) is used to design the classifier.

The probability of distortion d_m is given by

$$P(d_m \mid x_i) = p(x_i \mid d_m) / \sum_{j=1}^{N_d} p(x_i \mid d_j) \cdot P(d_j) \tag{G-43}$$

where N_d is the number of distortions and $P(d_j)$ is the *a priori* probability of distortion d_j.

An estimate of the opinion score probability function $P(\omega_q|x_i)$ can now be obtained. This is the probability of opinion score ω_q, where the classes (q) range from 1 (bad) to 5 (excellent). This function can be interpreted as the predicted frequency of listener panel scores corresponding to test speech parameter vector, x_i. The relationship is given by

$$P(\omega_q \mid x_i) = \sum_{m=1}^{N_d} m \; S(d_m \mid x_i) \cdot P(\omega_q \mid x_i, d_m) \tag{G-44}$$

where $S(\omega_q|x_i, d_m)$ is a histogram derived from subjective tests on distortion d_m and is the fraction of listener votes for quality score ω_q.

The predicted Mean Opinion Score can be found as follows:

$$MOS_i = \sum_q q \cdot P(\omega_q \mid x_i), (q = 1, \ldots, 5)$$
$$MOS = E\{MOS_i \mid x_i\} \tag{G-45}$$

Here, $E\{*\}$ is the expectation operator and can be approximated by averaging over all frames of the speech record.

Appendix I

(reference to 3.2.2)

Computer programs used

I.1 Computer program 1

```
10    PRINT "CALCULATION  OF  INFORMATION  INDEX  FOR  MNRU"
20    REM  New frequency weighting, Ti from BOSQUET
30    REM  PROGRAM  IIQCME.BAS, December 1988, written in MF  BASIC
40    INPUT "SYSTEM"; S$
50    INPUT "MOS"; Y$
60    DATA .05457, 4.1, .04733, 5.6, .06682, 6.4, .07497, 6.9, .06546, 7.4, .06622, 7.8, .05585, 8, .054, 8, .05273, 8.2,
      .05117, 8.2
70    DATA .04517, 8.2, .04706, 8.2, .05073, 8.2, .05561, 8.2, .0631, 8.2, .06886, 8.1
80    REM  Calculation for MNRU
90    FOR J = 1 TO 16
100   PRINT "Qseg over the band No"; J
110   INPUT  QS
120   READ  B, C
130   Q = QS + C
140   V = 3/(.1 + 10^(–Q/10))
150   I = B * V
160   II = II + I
170   NEXT  J
180   REM  Display of results
190   PRINT  S$, "II ="; II
200   LPRINT "          ";  S$; TAB(20); II; TAB(30); Y$; TAB(40)
210   END
```

I.2 Computer program 2

```
10    PRINT "CALCULATION  OF  INFORMATION  INDEX  FOR  CODECS"
20    REM  New frequency weighting, Ti from BOSQUET, revised equivalence with MNRU
30    REM  PROGRAM  IIQCDE.BAS, December 1988, written in MF  BASIC
40    INPUT "SYSTEM"; S$
```

```
50    INPUT "MOS"; Y$
60    K = 4.3429
70    DATA .05457, 4.1, .04733, 5.6, .06682, 6.4, .07497, 6.9, .06546, 7.4, .06622, 7.8, .05585, 8, .054, 8, .05273, 8.2,
      .05117, 8.2
80    DATA .04517, 8.2, .04706, 8.2, .05073, 8.2, .05561, 8.2, .0631, 8.2, .06886, 8.1
90    INPUT "QSEG over the band-QP = d (0 for PCM)"; SM
100   REM  Input of Qs in each band
110   FOR J = 1  TO  16
120   PRINT "Qseg over the band No"; J
130   INPUT  QS
140   READ  B, C
150   QC = QS + C
160   GOSUB  270
170   REM  Calculation of Information Index
180   V = 3/(.1 + 10^(–Q/10))
190   I = B * V
200   II = II + I
210   NEXT J
220   REM  Display of results
230   PRINT S$, "II ="; II
240   LPRINT "          ";  S$; TAB(20); SM; TAB(30); II; TAB(40); Y$; TAB(50)
250   END
260   REM  Calculation of equivalent Q
270   IF  QC > –3.57 THEN  QMC = K * LOG(EXP((QC + 5.15)/K) – 1)
280   IF  QC < –3.57 THEN  R2 = .15968 * QC – .71265 ELSE 320
290   D2 = (EXP (R2) – 1) / (EXP (R2) + 1)
300   Q = QC + D2 * SM
310   RETURN
320   IF  QC < 0 THEN  Q = QC + SM * (.276 * QC + .3859) ELSE  340
330   RETURN
340   R1 = .12543 * QC + .62185
350   D1 = (EXP (R1) = 1) / (EXP (R1) + 1)
360   Q = QMC + D1 * SM
370   RETURN
```

Appendix II

(reference to 3.2.2 and 3.3)

```
10   PRINT "Calculation of Information Index for NTT 600 sets (7 dB line)"
15   PRINT "with a mixture of white and exchange noise."
20   REM  Program  IIMNT6RE, written in MF  Basic, August 1990
30   INPUT "Room noise, dBA="; RN
40   INPUT "STMR, dB="; STMR
50   INPUT "Circuit noise level (dBm, sign changed) at input to 0 dB RLR end"
60   ICN0=–I
70   INPUT "Listening (L) or conversation (C) or terminate (T)"; A$
80   IF  A$="T" GOTO  640
90   IF  A$="C" GOTO  560
100  INPUT "Overall loudness rating (P79), dB="; OLR
110  LPRINT "   OLR="; OLR
120  GOSUB  730
130  REM  Correction for excessive loudness
140  IF  OLR>OPT  GOTO 380
150  X=2*OPT–OLR
160  GOTO  390
170  DIM  FE(20), CN(20), ST(20), EL(20), BKL(20), S(20), BJ(20), CJ(20), OLA(20)
180  DATA    –76.2,  –29.2,   –4.2,   32.4,   37.5,  56.4,  .00804,  0,      85.3
190  DATA    –28.9,    5.5,   –3.1,   31.2,   30.5,  61.5,  .01042,  1.25,   40.5
200  DATA    –15.8,    8.1,     .7,   29.5,   22.5,  62.9,  .0138,   2,      29.5
210  DATA     –2.6,   10.3,    5.5,   27.6,   13.4,  64.7,  .01788,  2.6,    18.4
220  DATA      9.2,   11.9,   12.6,   26.2,    7.4,  64.4,  .02392,  3.5,     8.4
230  DATA     16.8,   13.4,   16.3,   22.3,     .6,  61.1,  .03246,  4.9,     1.1
240  DATA     19.8,   13.2,   16.5,   22.7,   –3.7,  60.2,  .04471,  5.8,     –.9
250  DATA     18.9,   12,     13.4,   21.1,   –7.2,  59.8,  .05981,  6.35,   –1.1
260  DATA     15.7,   10.7,    8.8,   17.4,   –8.9,  56.7,  .07789,  7.25,   –1
270  DATA     14.7,    9.9,    5.9,    9.3,  –11.2,  52.8,  .0839,   7.35,   –1.2
280  DATA     14.5,    9.7,    4.8,    2.7,  –11.3,  48,    .0899,   7.8,    –1.6
290  DATA     13,      9.9,    3.5,    –.9,   –9.7,  45.6,  .09627,  8.05,   –1.5
300  DATA     13.8,    9.2,    4.8,   –7.1,   –8.1,  44.4,  .10376,  8.25,    –.9
310  DATA     13.6,    9.5,    5.4,  –12.4,   –5.4,  41.8,  .11097,  8.3,    –3.5
320  DATA      6.3,   11.3,   –1.8,  –20.4,   –4.1,  39.2,  .11859,  8.18,   –5.1
330  DATA     –5.3,    6.4,  –15.2,  –19.2,   –7.1,  35.1,  .12694,  7.95,     .8
340  DATA   –102.5,  –23,    –24.9,  –28.1,   –4.2,  31.4,  .13607,  7.57,   96.7
350  DATA   –263.6,  –64.5,  –40.5,  –38.4,   –2.5,  28.2,  .14506,  7.25,  259.2
360  DATA   –267.2,  –67.2,  –44.6,  –51.3,    1.4,  26.5,  .15487,  7.2,   263.3
370  DATA   –292.9,  –77.2,  –59.6,  –66.6,    6,    25.9,  .16554,  6.8,   281.3
```

```
380   X=OLR
390   DEF  FNP  (Y)=10^(Y/10)
400   IN=0
410   FOR  J=1 TO 20
420   READ  FE, CN, ST, EL, BKL, S, BJ, CJ, OLA
425   REM  Calculation and composition of signal to noise and equivalent ratio
430   PN=FNP(FE+RN–50–X+5)+FNP(CN+ICN0+60)+FNP(ST+RN–50–STMR+15)+FNP(EL+RN–50)
440   ZN=S–OLA–X+5–4.343*LOG(PN)
450   ZA=S–7.8–OLA–X–BKL
460   IF ZA>0  THEN  PE=(1+ZA/9.5)^2–1: GOTO 470
465   PE=10^(–10)
470   P=FNP (–ZN)+1/PE
480   Z=–4.343*LOG(P)
490   GOSUB  660
500   G=BJ*V
510   IN=IN+G
520   NEXT  J
530   PRINT “IN=”; IN; "OPT="; OPT
540   LPRINT “          RN(dBA)=”; RN; “STMR(dB)=”; STMR; “X(dB)=”; X; “ICN0(dB)=”; ICN0
545   LPRINT “          OPT="; OPT; "IN=”; IN
550   GOTO  70
560   RESTORE
570   REM  Speech power correction for sidetone and quality of conversation
580   IF  STMR>13  THEN  590  ELSE  610
590   CS=0
600   GOTO  620
610   CS=.3*(STMR–13)
620   X=X–CS+.4085*IN–9.87
630   GOTO  390
640   END
650   REM  Equivalence law and calculation of V
660   IF  Z<1.74 THEN  670  ELSE  690
670   Q=Z+CJ
680   GOTO  700
690   Q=.494*Z+.88+CJ
700   V=3/(.1+10^(–Q/10))
710   RETURN
720   REM  Determination of optimum OLR
730   IF  RN<30  THEN  DS=–2.4: GOTO  750
740   DS=.006*(RN–30)^2–2.4
750   RNS=RN–112.6+DS–STMR
760   RNL=RN–116
770   PC=10^(ICN0/10)
780   PRL=10^(RNL/10)
790   PRS=10^(RNS/10)
800   NT=4.343*LOG (PC+PRL+PRS+10^ (–8))
810   OPT=7.5–.14*(NT+80)
820   RETURN
```

Appendix III
(reference to 3.3.2)

Calculation of the DATA in Appendix II from primary data

III.1 Definition of the primary data

The primary data includes noise spectra and the electroacoustic characteristics of the type of telephone set used.

III.1.1 Noise spectra

Table III.1 gives the spectrum BDR of room noise (of the Hoth type) for which RN = 50 dBA. It also gives the spectra of circuit noises corresponding to ICNO = –60 dBmp; NDW corresponds to white noise, NDC to the mixed noise in the last column of Table C.1 (noise spectrum used by OPINE) lowered by 7.9 dB (since it corresponds to ICN = –56 dBmp and RLR = –3.9).

III.1.2 Electroacoustic characteristics

Those of the local systems, as defined in Recommendations P.64 and P.79, are:

S_{UMJ}	Sending sensitivity
S_{UJE}	Receiving sensitivity (measured on an artificial ear)
L_{MEST}	Sidetone path loss measured with an artificial mouth and ear
L_{RNST}	Sidetone path loss measured in a diffuse room noise field of 50 dBA, with an artifical ear.

Alternativity DSM = $L_{MEST} - L_{RNST}$ may be determined as the difference between the sending system sensitivity for a diffuse room noise field and S_{UMJ} (see 9/P.64).

The following chacteristics are also needed:

LE	Artificial/real ear correction (6/P.79)
L_{RNE}	Loss of the leak between the listener's ear and earcap, measured on a real ear.

The junction (representing the rest of the connection) is characterized by

D_1	Attenuation distortion
XL	Adjustable loss, independent of frequency.

NOTE – The information index is computed at the listener's end of the connection; a *d* is added for the characteristics at the distant (talker's) end.

III.2 Relations

III.2.1 **Transmission characteristics** (see 3.3.2.2 above)

First sending (SLR), receiving (RLR) loudness ratings and STMR are computed from the electroacoustic characteristics according to Recommendation P.79, as well as the value of XL which makes the overall loudness rating (OLR) equal to 5 dB.

At each frecuency, the overall loss between MRPd and an artificial ear at ERP, corresponding to OLR = 5 dB, is then

$$\mathrm{OLA} = -S_{UMJd} + \mathrm{D1d} + \mathrm{XLd} - S_{UJE}$$

III.2.2 **Effect of thresholds** (see 3.3.3.3)

From [24], clause V.2, in the absence of noise we have:

$$\mathrm{ZA} = \mathrm{S} - \mathrm{OLA} - (\mathrm{X} - 5) - \mathrm{LE} - (\mathrm{BK} + 12.8)$$

It is convenient to put BKL = BK + LE and we may write

$$\mathrm{ZA} = \mathrm{S} - \mathrm{OLA} - \mathrm{X} - \mathrm{BKL} - 7.8$$

as in line 450 of Appendix II.

III.2.3 **Noise components** (see 3.3.2.3)

These are computed by the following relations:

$$\begin{aligned}
\mathrm{FE} &= \mathrm{BDR} - \mathrm{OLA} + \mathrm{DSMd}\\
\mathrm{CN} &= \mathrm{ND} + S_{UJE} + 94\\
\mathrm{ST} &= \mathrm{BDR} - L_{RNST} - (15 - \mathrm{STMR})\\
\mathrm{EL} &= \mathrm{BDR} - L_{RNE}
\end{aligned}$$

III.3 Example

In addition to ND and BDR, Table III.1 gives the electroacoustic characteristics of a NTT 600 telephone set with 7 dB line.

NOTES

1 It seems that in Table D.1 (Local Telephone Circuit Sensitivity) L_{MESTi} was determined on an artificial ear and L_{RNSTi} with a room noise of 60 dBA on a real ear. In Table 4-4 L_{RNE} was obviously measured on a real ear. The both L_{MEST} and L_{RNE} were determined under conditions of high room noise where the listener presses the receiver tightly to his ear and the correction LE should not be applied for the calculation of ZN.

2 Since in this example we are considering symmetrical connections, the letter *d* is omitted. If RNd were different of RNn, a correction would be required in line 440 of Appendix II to the term including FE.

The calculations explained above give:

$$\mathrm{SLR} = 6.6 \qquad \mathrm{RLR} = -3.9 \qquad \mathrm{STMR} = 7 \qquad \mathrm{XL} = 1.3$$

and the data in line 180-370 of Appendix II.

TABLE III.1

Example of primary data

Frequency		Noise spectra		Electro-acoustic characteristics (NTT 600 sets, 7 dB line)						1 SRAEN filter
kHz	No.	NDC dB (1 V/Hz)	BDR dB (20 µPa/Hz)	S_{UMJ} dB (1 V/Pa)	S_{UJE} dB (1 Pa/V)	NDW dB (1 V/Hz)	L_{RNST} dB	L_{MEST} dB	L_{RNE} dB	D_1 dB
0.1	1	–83.15	32.43	–22.30	–40.00	–110.90	28.6	5.30	0.00	21.70
0.125	2	–85.83	31.23	–25.10	–2.70	–109.70	26.3	6.70	0.00	11.40
0.16	3	–88.37	29.53	–23.80	2.50	–108.00	20.8	5.00	0.00	6.90
0.2	4	–90.96	27.63	–18.80	7.30	–106.10	14.1	2.30	0.00	5.60
0.25	5	–93.36	26.23	–14.40	11.30	–103.50	5.6	–3.00	0.00	4.00
0.315	6	–95.19	23.03	–12.30	14.60	–100.70	–1.3	–6.40	0.70	2.10
0.4	7	–96.68	22.73	–12.50	15.90	–99.50	–1.8	–5.60	0.00	1.20
0.5	8	–97.66	21.13	–12.60	15.70	–98.90	–0.3	–3.60	0.00	0.70
0.63	9	–98.22	19.63	–12.30	14.90	–98.60	2.8	–2.10	2.20	0.30
0.8	10	–98.47	17.83	–11.90	14.40	–98.40	3.9	–0.40	8.50	0.00
1	11	–98.76	16.23	–11.60	14.50	–98.40	3.4	0.10	13.50	0.00
1.25	12	–98.91	14.63	–12.00	14.80	–98.50	3.1	0.00	15.50	0.00
1.6	13	–98.9	12.93	–12.00	14.10	–98.50	0.1	0.10	20.00	–0.10
2	14	–98.85	11.33	–9.80	14.40	–98.40	–2.1	–3.30	23.70	–0.20
2.5	15	–98.88	9.63	–10.00	16.20	–97.90	3.4	–5.00	30.00	–0.20
3.15	16	–99.09	7.83	–11.00	11.50	–101.20	15.0	2.70	27.00	0.00
4	17	–125.85	5.43	–16.80	8.90	–148.80	22.3	11.10	33.50	87.50
5	18	–128.45	2.63	–27.90	–30.00	–198.80	35.1	28.10	41.00	200
6.3	19	–131.15	–1.27	–32.00	–30.00	–198.80	35.3	32.70	50.00	200
8	20	–131.15	–6.57	–40.00	–40.00	–198.80	45.0	40.00	60.00	200

References

[1] CAVANAUGH (J. R.), HATCH (R. W.) and SULLIVAN (J. L.): Models for the subjective effects of loss, noise and talker echo on telephone connections, *B.S.T.J.*, Vol. 55, No. 9, pp. 1319-1371, November, 1976.

[2] SULLIVAN (J. L.): A laboratory system for measuring loudness loss of telephone connections, *B.S.T.J.*, Vol. 50, No. 8, pp. 2663-2739, October 1971.

[3] CAVANAUGH (J. R.), HATCH (R. W.) and NEIGH (J. L.): A model for the subjective effects of listener echo on telephone connections, *B.S.T.J.*, Vol. 59, No. 6, pp. 1009-1060, July-August 1980.

[4] CCITT – Contribution COM XII-No. 13, (Bell-Northern Research), Study Period 1981-1984, Geneva, 1981.

[5] CCITT – Contribution COM XII-No. 14, (Bell-Northern Research), Study Period 1981-1984, Geneva, 1981.

[6] DAUMER (W. R.) and CAVANAUGH (J. R.): A subjective comparison of selected digital codecs for speech, *B.S.T.J.*, Vol. 57, No. 9, pp. 3119-3165, November 1978.

[7] CCITT – Contribution COM XII-No. 173, (American Telephone and Telegraph Company), Study Period 1977-1980, Geneva, 1979.

[8] CCITT – Contribution COM XII-No. 94, (American Telephone and Telegraph Company), Study Period 1981-1984, Geneva, 1982.

[9] CCITT – Contribution COM XII-No. 159, (American Telephone and Telegraph Company), Study Period 1977-1980, Geneva, 1979.

[10] CCITT – Contribution COM XII-No. 158, (American Telephone and Telegraph Company), Study Period 1981-1984, Geneva, 1983.

[11] AHERN (W. C.), DUFFY (F. P.) and MAHER (J. A.): Speech signal power in the switched message network, *B.S.T.J.*, Vol. 57, No. 7, pp. 2695-2726, September 1978.

[12] CUMMISKEY (P.), JAYANT (N. S.) and FLANAGAN (J. L.): Adaptive quantization in differential PCM coding of speech, *B.S.T.J.*, Vol. 52, No. 7, pp. 1105-1118, September 1973.

[13] RICHARDS (D. L.): Calculation of opinion scores for telephone connections, *Proc. I.E.E.*, Vol. 121, No. 5, pp. 313-323, May 1974.

[14] CCITT Contribution COM XII-No. 113, Some remarks on the structure of a telephone connection assessment model, Study Period 1977-1980, Geneva, October 1978.

[15] CCITT Contribution COM XII-No. 129, Use of a telephone connection assessment model in the study of Question 15/XII, Study Period 1977-1980, Geneva, February 1979.

[16] WEBB (P. K.): The background and philosophy of the telephone network assessment program (CATNAP), British Post Office Research Department Report No. 752, 1979.

[17] *Prediction of transmission qualities from objective measurements*, Yellow Book, Volume V, Supplement No. 4, ITU, Geneva, 1981.

[18] RICHARDS (D. L.): Telecommunication by speech: The transmission performance of telephone networks, Chapter 3, *Butterworths*, London, 1973.

[19] RICHARDS (D. L.): Telecommunication by speech: The transmission performance of telephone networks, Chapter 2, *Butterworths*, London, 1973.

[20] RICHARDS (D. L.): Telecommunication by speech: The transmission performance of telephone networks, p. 57, *Butterworths*, London, 1973.

[21] RICHARDS (D. L.): Transmission performance of telephone networks containing P.C.M. links, *Proceedings of the I.E.E.*, Vol. 115, No. 9, pp. 1245-1258, September 1968.

[22] CCITT Handbook *Transmission planning of switched telephone network*, Chapter II, Annex 3, ITU, Geneva, 1976.

[23] BOERYD (A.): Subscriber reaction due to unbalanced transmission levels, *Third International Symposium on Human Factors in Telephony*, 1966, pp. 39-43, The Hague, 1967.

[24] LALOU (J.): The information index: an objective measure of speech transmission performance. Annales des Télécommunications, Vol. 45, No. 1-2, pp. 47-65, 1990.

[25] RICHARDS (D. L.): Calculation of opinion scores for telephone connections, *Proc. I.E.E.*, Vol. 121, No. 5, pp. 313-323, May 1974.

[26] ALCAIM (A.): Essai de détermination d'un indice objectif de mesure de la qualité des codeurs. *CNET*, Report RP/LAA/TSS/208, May 1984.

[27] RICHARDS (D. L.): private communication

[28] OPINE (Rev. 2.0.), Electrical Communication Laboratories NTT, September 1986.

[29] RICHARDS (D. L.), BARNES (G. J.): Pay-off between quantizing distortion and injected circuit noise, in *Proc. ICASSP 82* Vol. 2, pp. 984-987, Paris, May 1982.

[30] CCITT – Contribution COM XII-No. 174, Transmission performance objective evaluation model for fundamental factors, (NTT), Geneva, 1983.

[31] CCITT – Contribution COM XII-No. 235, Calculation method of OPINE, (NTT), Geneva 1984.

[32] CCITT – Contribution COM XII-No. 10, Objective evaluation model of telephone transmission performance for fundamental transmission factors and quantizing distortion, (NTT), Geneva 1985.

[33] OSAKA (N.) and KAKEHI (K.): Objective model for evaluating telephone transmission performance, *Review of ECL*, Vol. 34, No. 4, 1986.

[34] IAI (S.) and IRII (H.): Subjective assessment of echo delay time effect, *Conference record of Acous. Soc. of Japan*, 2-7-8, (1983-03) (in Japanese).

[35] IAI (S.) *et al.*: A study on subjective assessment of telephone speech sidetone *Conference record of Acous. Soc. of Japan*, 2-2-3, (1978-05) (in Japanese).

[36] OSAKA (N.) and KAKEHI (K.): A study on the psychological factors that affect the opinion evaluation of telephone transmission performance., *Trans. IECE*, Vol. J69-A, No. 5, 1986 (in Japanese).

[37] IRII (H.): Loudness equivalent attenuation of speech in the presence of noise, *Conference Record of Acous. Soc. of Japan*, 1-4-6 (1975-05) (in Japanese).

[38] ISO Recommendation R226: Normal equal loudness contours for pure tones and normal threshold of hearing under free-field listening conditions, Dec. 1961.

[39] CAVANAUGH, (J. R.), HATCH, (R. W.) and SULLIVAN (J. L.): Models for the subjective effects of loss, noise and talker echo on telephone connections, *B.S.T.I.* Vol. 55, No. 9, 1976.

[40] CCITT Recommendation *Measurement of the AEN Value of a commercial telephone system*. Yellow Book, Vol. V, Rec. P.45, ITU, Geneva, 1981.

[41] CCITT – Contribution COM XII-No. 175, BNR, Canada, 1987.

[42] CCITT – Delayed Contribution D.79, USA, 1990.

[43] CCITT – Contribution COM XII-No. 114, Pays-Bas, 1991.

Printed in Switzerland
Geneva, 1994